

## **General Disclaimer**

### **One or more of the Following Statements may affect this Document**

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

# AgRISTARS

SR-63-0444

A Joint Program for  
Agriculture and  
Resources Inventory  
Surveys Through  
Aerospace  
Remote Sensing

## Supporting Research

AUGUST 1983

FINAL REPORT -- CENTER OF EXCELLENCE FOR APPLIED MATHEMATICAL  
AND STATISTICAL RESEARCH IN SUPPORT OF DEVELOPMENT OF MULTICROP  
PRODUCTION MONITORING CAPABILITY

WAYNE A. WOODWARD

HENRY L. GRAY



CENTER FOR APPLIED MATHEMATICAL AND STATISTICAL RESEARCH  
SOUTHERN METHODIST UNIVERSITY  
DALLAS, TEXAS 75275



(E84-10062) CENTER OF EXCELLENCE FOR  
APPLIED MATHEMATICAL AND STATISTICAL  
RESEARCH IN SUPPORT OF DEVELOPMENT OF  
MULTICROP PRODUCTION MONITORING CAPABILITY  
Final Report, 1 Feb. - (Southern Methodist

N84-19953

Unclass

G3/43 00062

1. Report No. SR-63-04444		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle FINAL REPORT -- Center of Excellence for Applied Mathematical and Statistical Research in Support of Development of Multicrop Production Monitoring Capability				5. Report Date August 31, 1983	
				6. Performing Organization Code	
7. Author(s) Wayne A. Woodward, H. L. Gray				8. Performing Organization Report No.	
9. Performing Organization Name and Address Center for Applied Mathematical and Statistical Research Southern Methodist University Dallas, Texas 75275				10. Work Unit No.	
				11. Contract or Grant No. NAS 9-16438	
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Johnson Space Center Houston, Texas Tech. Monitor: G. Houston				13. Type of Report and Period Covered Final Report Feb. 1-1983-August 31, 1983	
				14. Sponsoring Agency Code	
15. Supplementary Notes					
16. Abstract  <p>In this report we describe the research accomplishments of the Center of Excellence for Applied Mathematical and Statistical Research (CAMSR) at Southern Methodist University in support of contract NAS 9-16438 during the period of February 1, 1983 to August 31, 1983. Our research efforts have been in support of the development of multicrop production monitoring capability. In particular we have investigated segment level proportion estimation techniques based upon a mixture model. Our recent efforts have dealt primarily with the evaluation of current techniques and the development of alternative ones. A comparison of techniques is provided in this report on both simulated and LANDSAT data along with an analysis of the quality of profile variables obtained from LANDSAT data.</p> <p style="text-align: right;"><b>"Made available under NASA sponsorship in the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof"</b></p>					
17. Key Words (Suggested by Author(s)) Final Report Center of Excellence Proportion estimation minimum distance estimation maximum likelihood estimation LANDSAT Data				18. Distribution Statement	
19. Security Classif. (of this report) U		20. Security Classif. (of this page) U		21. No. of Pages 97	
22. Price*					

**FINAL REPORT**

**Center of Excellence for Applied Mathematical and  
Statistical Research in Support of Development  
of Multicrop Production Monitoring Capability**

**Wayne A. Woodward and H. L. Gray**

**Center for Applied Mathematical and Statistical Research  
Southern Methodist University  
Dallas, Texas 75275**

**August 1983**

## TABLE OF CONTENTS

	Page
INTRODUCTION . . . . .	1
(1) Normal-based MDE vs MLE. . . . .	3
(2) Minimum Hellinger Distance Estimation. . . . .	5
(3) Weibull Based MD Estimation. . . . .	6
(4) LANDSAT Data . . . . .	13

Appendix A: Minimum Distance Estimation of Mixture  
Model Parameters -- Asymptotic Results and Simulation  
Comparisons with Maximum Likelihood  
(SR-63-04427)

Appendix B: Minimum Hellinger Distance  
Estimation of Mixture Model Parameters  
(SR-63-04433)

Appendix C: Proportion Estimation in Mixtures  
of Asymptotic Distributions  
(SR-63-04409)

## FINAL REPORT

### Center of Excellence for Applied Mathematical and Statistical Research in Support of Development of Multicrop Production Monitoring Capability

August 31, 1983

#### Introduction

In this report we summarize the efforts undertaken by the Center for Applied Mathematical and Statistical Research at Southern Methodist University in support of the contract NAS 9-16438 since January 31, 1983. For a discussion of the progress made on this contract prior to January 31, 1983, reference should be made to Final Report SR-63-04408. Our recent efforts have dealt primarily with an evaluation of current techniques for mixture model proportion estimation along with investigation into alternative techniques. Mixture modeling procedures currently utilized by NASA, e.g. CLASSY, assume a mixture of normal components. In addition, associated parameter estimation is accomplished using maximum likelihood (ML) methods based on the normality assumption since these ML estimators are optimal when the normality assumption is valid. However, it is well known that ML estimation procedures are highly sensitive to violations of the underlying assumptions. Recent implementation of the mixture model has involved use of

feature variables from the Badhwar profile model. In particular, the feature variables currently in use are  $T_p$  - time of peak greenness,  $G(T_p)$  - peak greenness, and  $\nabla$  - a measure of the length of the growing season. The normality of these feature variables has been an issue of recent concern.

Our results and investigations can be grouped into four major categories:

- (1) Further results on the comparison of normal based MLE and MDE (Cramér-von Mises distance)
- (2) Use of the Hellinger metric as an alternative to the Cramér-von Mises distance used in calculating the MDE
- (3) Investigation into the use of the Weibull, as an alternative to the normal for modeling component distributions.
- (4) Implementation of the estimation procedures on LANDSAT data in an effort to:
  - (a) investigate the normality (or non-normality) of the Badhwar feature variables.
  - (b) compare the performance of the MLE and MDE in a "real data" situation.

The progress which has been made in these areas is discussed in this report.

### (1) Normal-based MDE vs MLE

One of our primary investigations has been the comparison of normal based MD and ML estimation of the mixture proportion for simulated two component mixture models. We have compared the estimation procedures for simulated mixtures of normal and of non-normal components. Our investigations in this area were previously documented in NASA technical report SR-62-04376. In that report we showed ML procedures to be superior when the normal component assumption is valid while MD procedures perform better on the simulated mixtures of components which represent symmetric departures from normality. Mixtures of  $t(4)$  components were examined in that report.

Our recent results have included more extensive simulations in which double exponential and  $t(2)$  components were examined. The double exponential was chosen since it has lighter tails than a  $t(4)$  yet heavier than a normal distribution. Tests for goodness-of-fit usually have little power in distinguishing normal and double exponential data. In addition,  $t(2)$  components were examined in order to compare the estimation procedures in a heavier tailed setting than the  $t(4)$ . In fact, the  $t(2)$  distribution has infinite variance, and not surprisingly, realizations often



have a few extreme observations. Our results show that in both of these non-normal situations, the MDE provides better proportion estimates than the MLE. This improvement is particularly striking for the  $t(2)$  simulations as the MDE seems to be relatively insensitive to a few extreme values.

In these simulations we have initiated the iterative routines used to calculate the MDE and the MLE with starting values obtained using a somewhat ad-hoc quasi-clustering technique. We have observed that in some situations, particularly those with heavy overlap between component distributions, the starting values perform better than both the MLE and MDE, an interesting finding since the starting value routine is easy to implement and is very fast since it does not involve iteration.

Asymptotic results have been obtained which establish the strong consistency and asymptotic normality of the MDE in the mixture-of-normals setting. The form of the asymptotic variance of the MDE is available from these results so that the asymptotic relative efficiencies (AREs) of the MDE relative to the MLE can be found. We have calculated these AREs for several parameter configurations. These AREs are fairly comparable to the empirical finite sample results. The following reports have been written since January 31, 1983 concerning our work in this area. Report [1] is included as Appendix A in this document.

- [1] "Minimum Distance Estimation of Mixture Model Parameters - Asymptotic Results and Simulation Comparisons with Maximum Likelihood" by Wayne A. Woodward, William C. Parr, William R. Schucany, and Henry L. Gray (SR-63-04427), June 1983.
- [2] "A Comparison of Minimum Distance and Maximum Likelihood Estimation of a Mixture Proportion" by Wayne A. Woodward, William C. Parr, William R. Schucany, and Hildegard Lindsey - submitted to Journal of the American Statistical Association.

#### (2) Minimum Hellinger Distance Estimation

We have also investigated the use of the Hellinger metric for calculating the MDE. In our previous work, the Cramér-von Mises distance has been used exclusively for this calculation. The minimum Hellinger distance estimator (MHDE) is of interest to aerospace remote sensing since it has the potential of providing robust proportion estimates under deviations from normality while maintaining performance comparable to the MLE when the underlying components actually are normal. However, our initial implementations of this procedure have shown that although the results are encouraging, the iterative procedure is highly sensitive to

starting values and is computationally more difficult than the MDE based upon Cramér-von Mises distance. More investigation is needed before the MHDE can be considered to be a viable alternative for proportion estimation.

As a result of our investigations of the MHDE, the following report has been written and is included here as Appendix B.

[3] "Minimum Hellinger Distance Estimation of Mixture Model Parameters" by Wayne A. Woodward and Paul W. Eslinger (SR-63-04433), July 1983.

### (3) Weibull Based MD Estimation

The MD and ML estimation procedures discussed in the previous sections were both based upon a mixture of normal components. Our results showed that although the MDE is more robust to symmetric departures from component normality, neither normal based procedure provided adequate results in the presence of asymmetric departures. We have investigated the use of the Weibull distribution as an alternative to the normal since the Weibull can be symmetric or skewed (to the right or to the left), and it therefore provides a very flexible model. The density function for the three parameter Weibull is given by

$$f(x) = \frac{\gamma}{\beta} \left(\frac{x-\alpha}{\beta}\right)^{\gamma-1} e^{-\left(\frac{x-\alpha}{\beta}\right)^{\gamma}}, \quad x \geq \alpha \quad (1)$$

where  $\beta > 0$  and  $\gamma > 0$ . The mean and variance are given by

$$\begin{aligned} \mu &= \alpha + \beta \Gamma\left(\frac{1}{\gamma} + 1\right) \\ \sigma^2 &= \beta^2 \left\{ \Gamma\left(\frac{2}{\gamma} + 1\right) - \Gamma^2\left(\frac{1}{\gamma} + 1\right) \right\}. \end{aligned} \quad (2)$$

The parameter  $\gamma$  serves as a shape parameter. When  $\gamma=3.6$  the Weibull is symmetric and in fact, quite similar to the normal distribution. The Weibull is skewed to the left or to the right depending on whether  $\gamma > 3.6$  or  $\gamma < 3.6$  respectively. The following technical report addresses the use of the Weibull in mixture proportion estimation. It is included here as Appendix C.

- [4] "Proportion Estimation in Mixtures of Asymmetric Distributions" - Wayne A. Woodward, Richard F. Gunst, Hildegard Lindsey, and H.L. Gray.  
(SR-63-04409), May 1983.

The  $\chi^2(9)$  was used in the simulations in report [4] to assess the effect of component asymmetry. In that report, the iterative procedures were started at "truth" rather than at starting values obtained from the data. In Table 1 we present the results of a recent and more extensive set of

simulations than those quoted in [4]. In the new simulations, starting values for  $\mu_i$ ,  $\sigma_i^2$ , and  $p$ ,  $i=1,2$  were obtained from the data as discussed in [2]. The starting values for  $\mu_i$  and  $\sigma_i^2$  were then converted to estimates of  $\alpha_i$  and  $\beta_i$  using equations (2) with  $\gamma_i=3.6$ ,  $i=1,2$ . The starting value estimate for  $p$  remains unchanged. The simulations summarized in Table 1 are based on simulated mixtures of normal components while those in Table 2 are based on simulated mixtures of  $\chi^2(9)$  distributions. As in [4] we examined overlaps, as defined in [2], of .10 and .03, and mixing proportions of .25, .50, and .75. We have added the case in which the variance of component 1 is twice that of component 2. In this table we compare the MLE based on a mixture of normals model (MLEN), the MDE based on a mixture of normals model (MDEN), and the MDE based on a mixture of Weibulls model (MDEW).

The results here are similar to those shown in SR-63-04409 in that the normal based procedures performed better on the mixtures of normal components while the Weibull based MDE was generally superior on mixtures of  $\chi^2(9)$  components. Again, the starting value routine obtained estimators which were competitive with and often better than those estimators obtained through the iterative routines.

A few other comments are in order here. First, we believe that if the asymmetry can be assumed to be in only one direction (probably to the right for the profile

Table 1. Comparison of Proportion Estimation Techniques  
 Simulated Mixtures of Normal Components  
 $n = 200$   
 number of repetitions = 100

$$\sigma_1^2 = \sigma_2^2$$

		<u>Overlap = .10</u>			<u>Overlap = .03</u>		
		$\hat{p}$	$\hat{Bias}$	$\hat{MSE}$	$\hat{p}$	$\hat{Bias}$	$\hat{MSE}$
$p = .25$	MLEN	.27	.02	.023	.26	.01	.002
	MDEN	.31	.06	.045	.27	.02	.005
	MDEW	.33	.08	.042	.33	.08	.015
	Starts	.30	.05	.009	.29	.04	.005
$p = .50$	MLEN	.50	.00	.020	.49	-.01	.002
	MDEN	.50	.00	.019	.49	-.01	.002
	MDEW	.49	-.01	.028	.49	-.01	.005
	Starts	.51	.01	.007	.50	.00	.005

$$\sigma_1^2 = 2\sigma_2^2$$

		<u>Overlap = .10</u>			<u>Overlap = .03</u>		
		$\hat{p}$	$\hat{Bias}$	$\hat{MSE}$	$\hat{p}$	$\hat{Bias}$	$\hat{MSE}$
$p = .25$	MLEN	.24	-.01	.011	.24	-.01	.002
	MDEN	.31	.06	.031	.25	.00	.005
	MDEW	.36	.11	.082	.25	.00	.011
	Starts	.23	-.02	.007	.25	.00	.003
$p = .50$	MLEN	.49	-.01	.012	.50	.00	.002
	MDEN	.50	.00	.016	.50	.00	.002
	MDEW	.49	-.01	.027	.50	.00	.007
	Starts	.41	-.09	.015	.45	-.05	.007
$p = .75$	MLEN	.70	-.05	.025	.74	-.01	.002
	MDEN	.64	-.11	.057	.73	-.02	.004
	MDEW	.59	-.16	.057	.71	-.04	.009
	Starts	.59	-.16	.035	.66	-.09	.012

Table 2. Comparison of Proportion Estimation Techniques  
Simulated Mixtures of  $\chi^2(9)$  Components

$n = 200$

number of repetitions = 100

$$\sigma_1^2 = \sigma_2^2$$

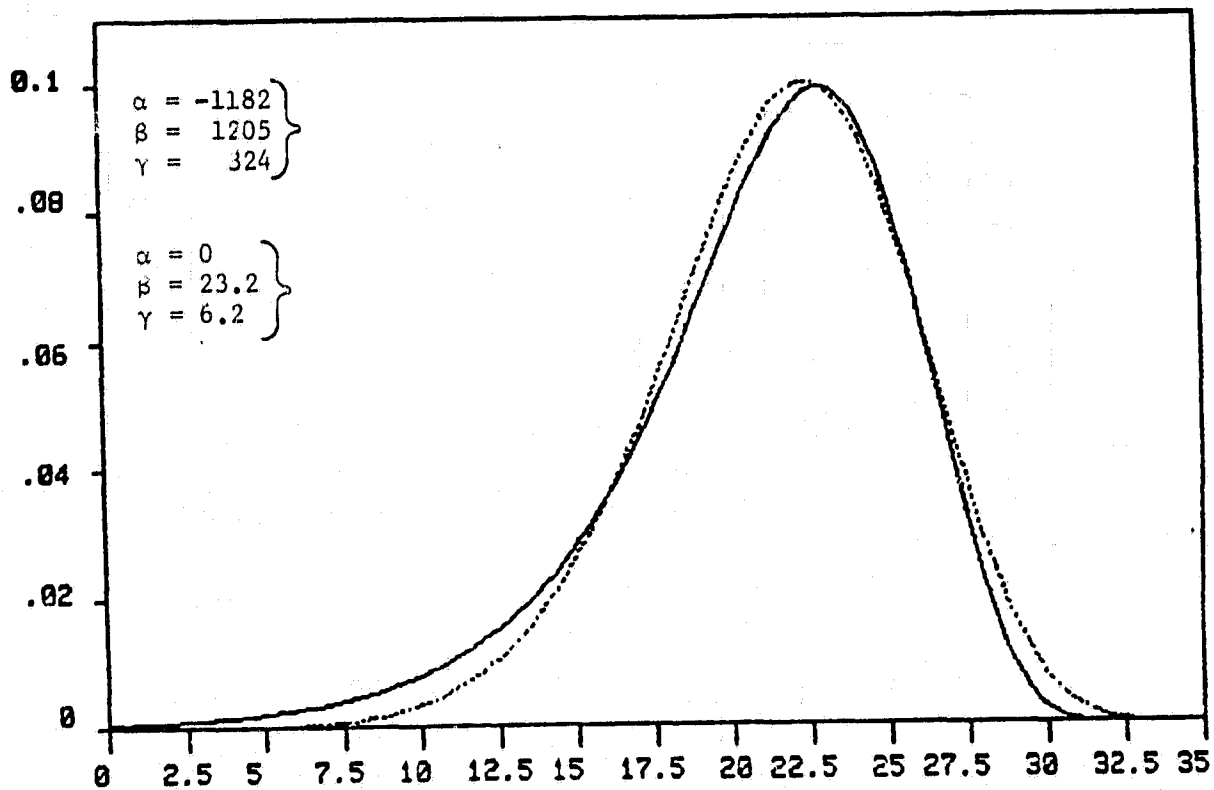
		<u>Overlap = .10</u>			<u>Overlap = .03</u>		
		$\hat{p}$	$\hat{Bias}$	$\hat{MSE}$	$\hat{p}$	$\hat{Bias}$	$\hat{MSE}$
$p = .25$	MLEN	.27	.02	.096	.17	-.08	.007
	MDEN	.34	.09	.106	.17	-.08	.008
	MDEW	.35	.10	.049	.32	.06	.011
	Starts	.32	.07	.035	.27	.02	.003
$p = .50$	MLEN	.26	-.24	.062	.41	-.09	.011
	MDEN	.29	-.21	.058	.42	-.08	.009
	MDEW	.42	-.08	.023	.49	-.01	.005
	Starts	.48	-.02	.007	.51	.01	.004
$p = .75$	MLEN	.48	-.27	.080	.65	-.10	.013
	MDEN	.46	-.29	.095	.63	-.12	.016
	MDEW	.55	-.20	.075	.71	-.04	.006
	Starts	.67	-.08	.013	.66	-.09	.013

		<u>Overlap = .10</u>			<u>Overlap = .03</u>		
		$\hat{p}$	$\hat{Bias}$	$\hat{MSE}$	$\hat{p}$	$\hat{Bias}$	$\hat{MSE}$
$p = .25$	MLEN	.16	-.09	.050	.19	-.06	.006
	MDEN	.28	.03	.069	.18	-.07	.008
	MDEW	.37	.12	.062	.31	.06	.010
	Starts	.26	.01	.026	.25	.00	.003
$p = .50$	MLEN	.28	-.22	.053	.43	-.07	.007
	MDEN	.31	-.19	.041	.43	-.07	.007
	MDEW	.45	-.05	.015	.49	-.01	.003
	Starts	.41	-.09	.015	.45	-.05	.007
$p = .75$	MLEN	.46	-.29	.089	.65	-.10	.012
	MDEN	.50	-.25	.076	.64	-.11	.016
	MDEW	.58	-.17	.049	.69	-.06	.008
	Starts	.60	-.15	.030	.61	-.14	.023

variables under consideration), then the estimation results shown in Tables 1 and 2 can be improved. Another interesting finding was made during these simulations concerning the 3-parameter Weibull. Although we only show summary results in this report, parameter estimates for all 7 of the parameters of the fitted mixture-of-Weibulls is printed out by the simulation program for each sample generated. For several samples  $\alpha(<0)$  and  $\beta$  were very large in absolute value, sometimes greater than 1000. These parameter values were associated with a  $\gamma$  smaller in magnitude than  $\alpha$  and  $\beta$  but substantially larger than 3.6. Although these parameter values appear to be "very bad", plots of the associated 3-parameter Weibull densities showed to be consistent with the data with only very small probability being associated with the interval between  $\alpha$  and 0. In Figure 1 we show two 3-parameter Weibull densities, one associated with parameters  $\alpha = -1182$ ,  $\beta = 1205$ , and  $\gamma = 324$  while the other density has parameters  $\alpha = 0$ ,  $\beta = 21.2$ ,  $\gamma = 5.8$ . We see that the densities are very similar although the parameter values differ dramatically. Thus, the 3-parameter Weibull seems to suffer from a "practical non-identifiability" which may or may not be a problem in our setting. If only proportion estimates are desired, then this lack of "identifiability" of the component Weibulls may not cause difficulties. It is clear that the component Weibull parameter estimates can be very misleading.



Figure 1. Practical Non-Identifiability of the 3-Parameter Weibull



#### (4) LANDSAT Data

The relative performance of the MDE and MLE has been examined through extensive simulation investigations. These investigations have been an important first step in understanding the behavior of the estimators in controlled normal and non-normal mixtures. It has been shown that lack of component normality can severely degrade the performance of the MLE. In fact, we have seen that mild non-normality (double exponential) can cause the "optimal" MLE to perform in a less than optimal manner. Another key concern involves the symmetry of component distributions since our simulations have shown that such skewness can have adverse effects on normal based procedures.

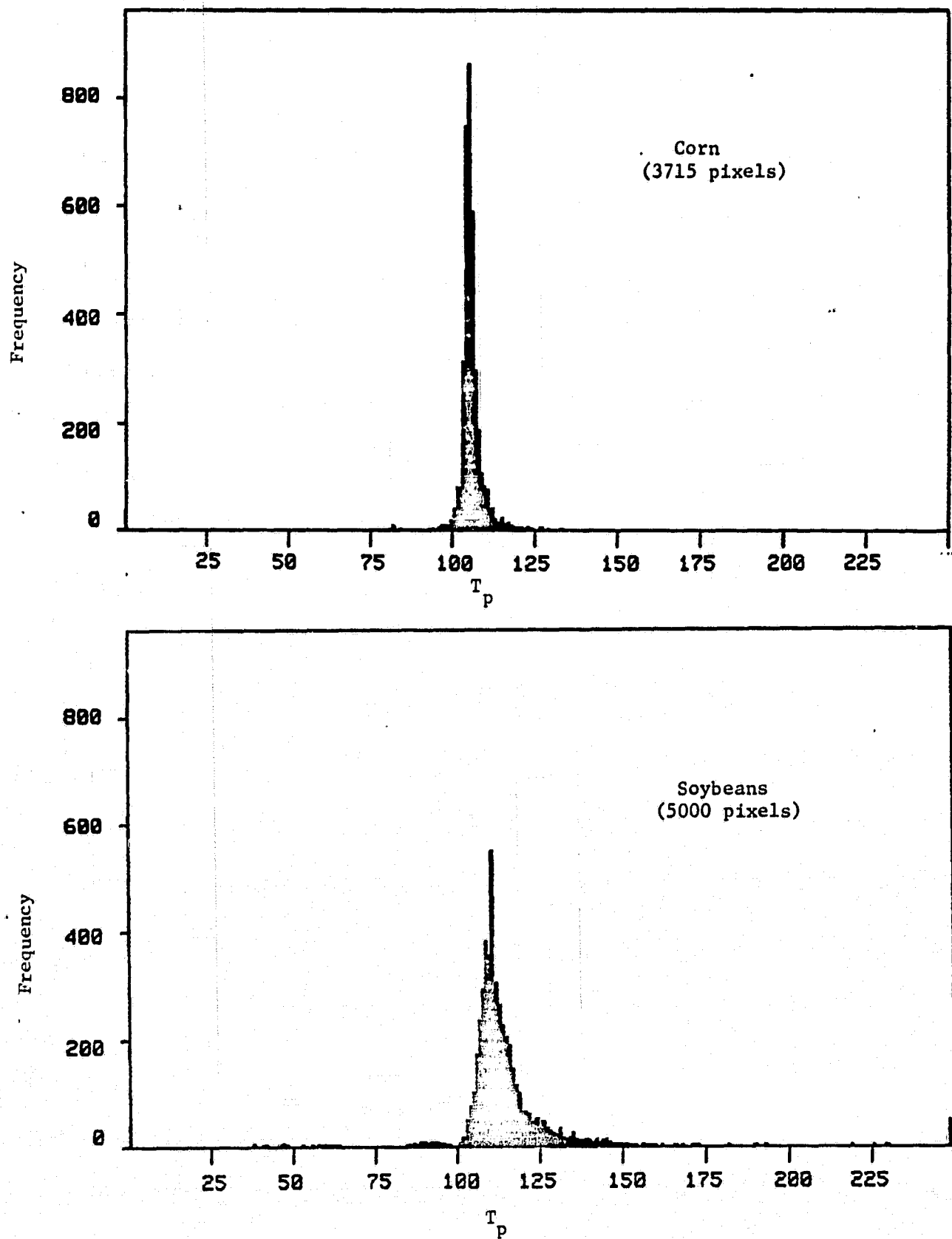
The performance of the estimation schemes on LANDSAT data is, of course, of ultimate importance. The key questions which are of interest in this respect are:

- (a) Are the feature variables from the Badhwar profile model normal? If not, what type of non-normality is encountered?
- (b) How do the estimation procedures compare on this data?

In an attempt to provide answers to these questions, we

have utilized data from the Fundamental Research Data Base. This data base consists of eighteen segments on which ground truth and the Badhwar feature variables are available for each pixel. In our investigations we identified the pure pixels on each segment and related these back to their ground truth labels. Our simulation investigations have been based on mixtures of two univariate component distributions. Therefore, the current interest concerns the ability of the estimation procedures to estimate crop proportions in this univariate, two component, real data setting. Accordingly, we identified "pairs" of crops from these 18 segments for which proportion estimation would be useful. That is, from a given segment we identified two crops, say corn and soybeans, and considered the related pixels to constitute a mixture population. In an attempt to further understand the data for these mixture populations, histograms of the component distributions and of the mixture distribution were drawn for each of the three feature variables  $T_p$ ,  $\nabla$ , and  $G(T_p)$ . In Figures 2 - 7 we display these histograms for the corn and soybean pure pixels of Segment 1380, a 1978 Minnesota segment. Several observations can be made concerning the histograms. First, there is clear visual separation between corn and soybeans on the basis of  $G(T_p)$ , a small amount of separation on  $T_p$ , and no separation on  $\nabla$ . Notice that what appears to be a second peak in the mixture model for  $\nabla$  in Figure 5 appears as a "spurious" peak in the

Figure 2. Histograms of  $T_p$  for Corn and Soybean Component  
Distributions based upon Pure Pixels from Segment 1380



ORIGINAL PAGE IS  
OF POOR QUALITY

Figure 3. Mixture Histogram based on  $T_p$  for Corn and Soybean Components of Figure 2

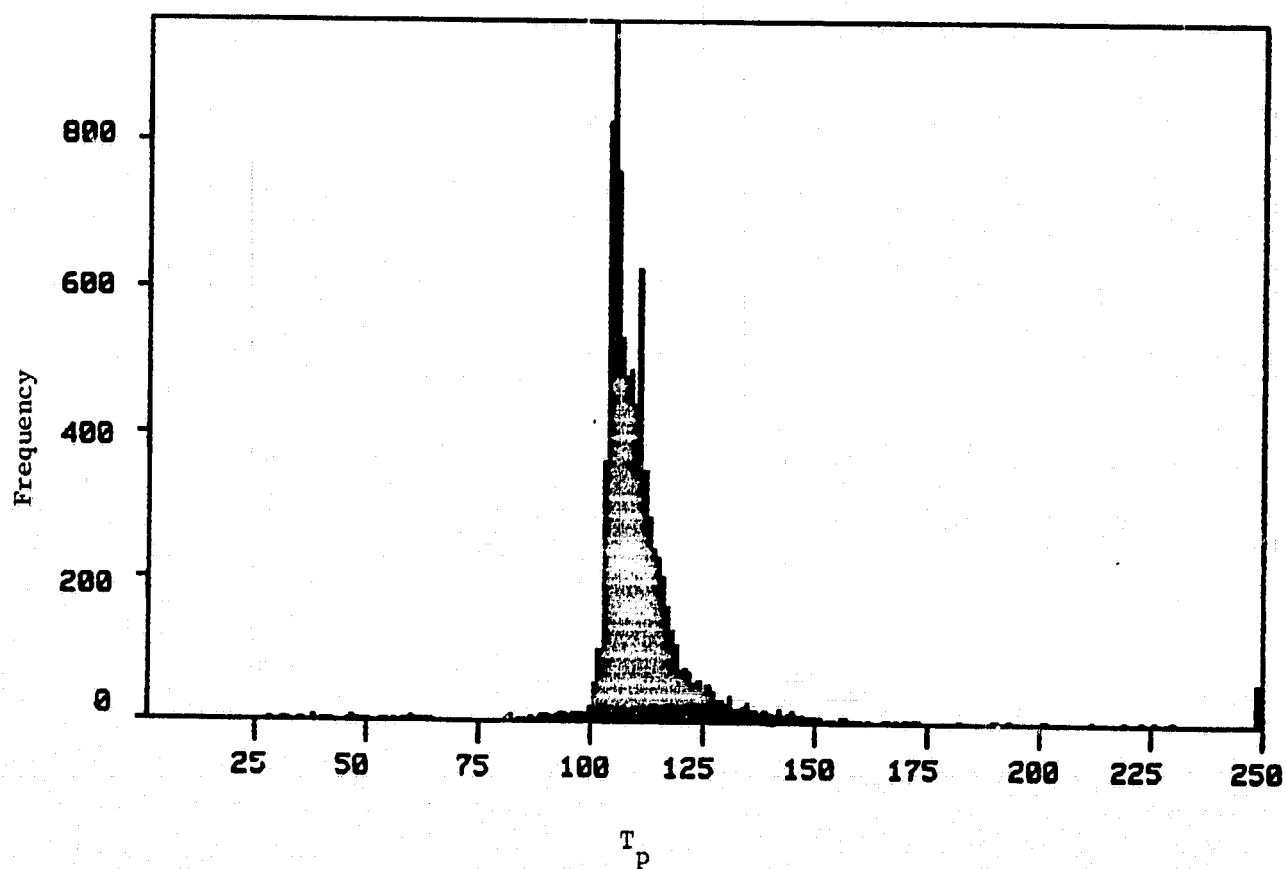
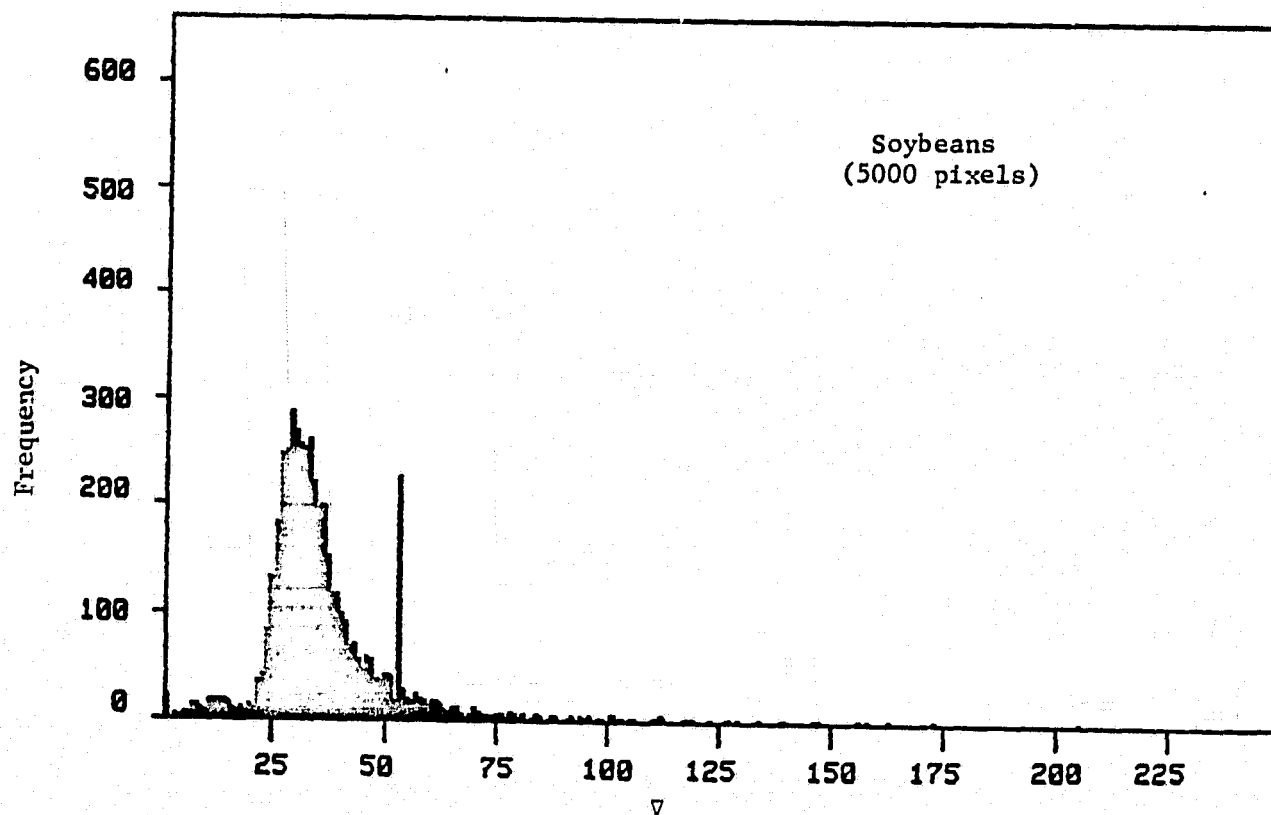
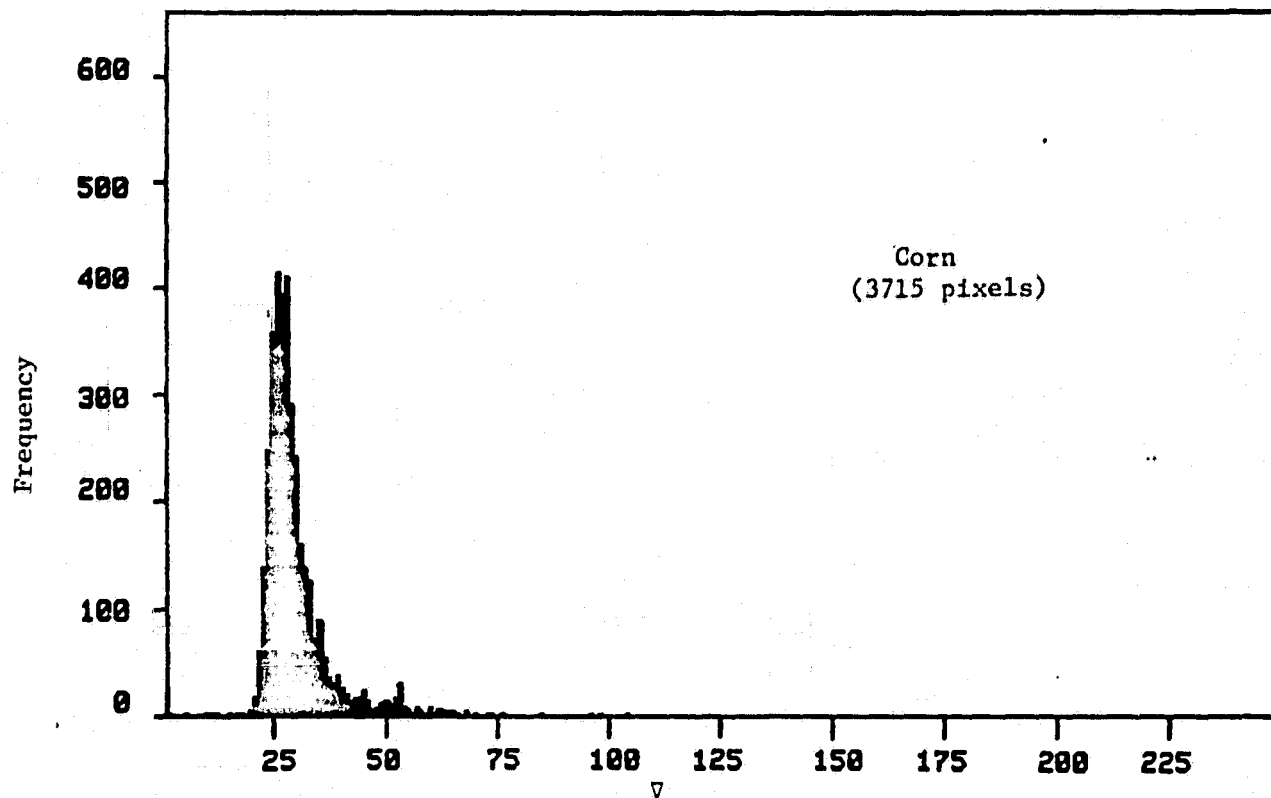
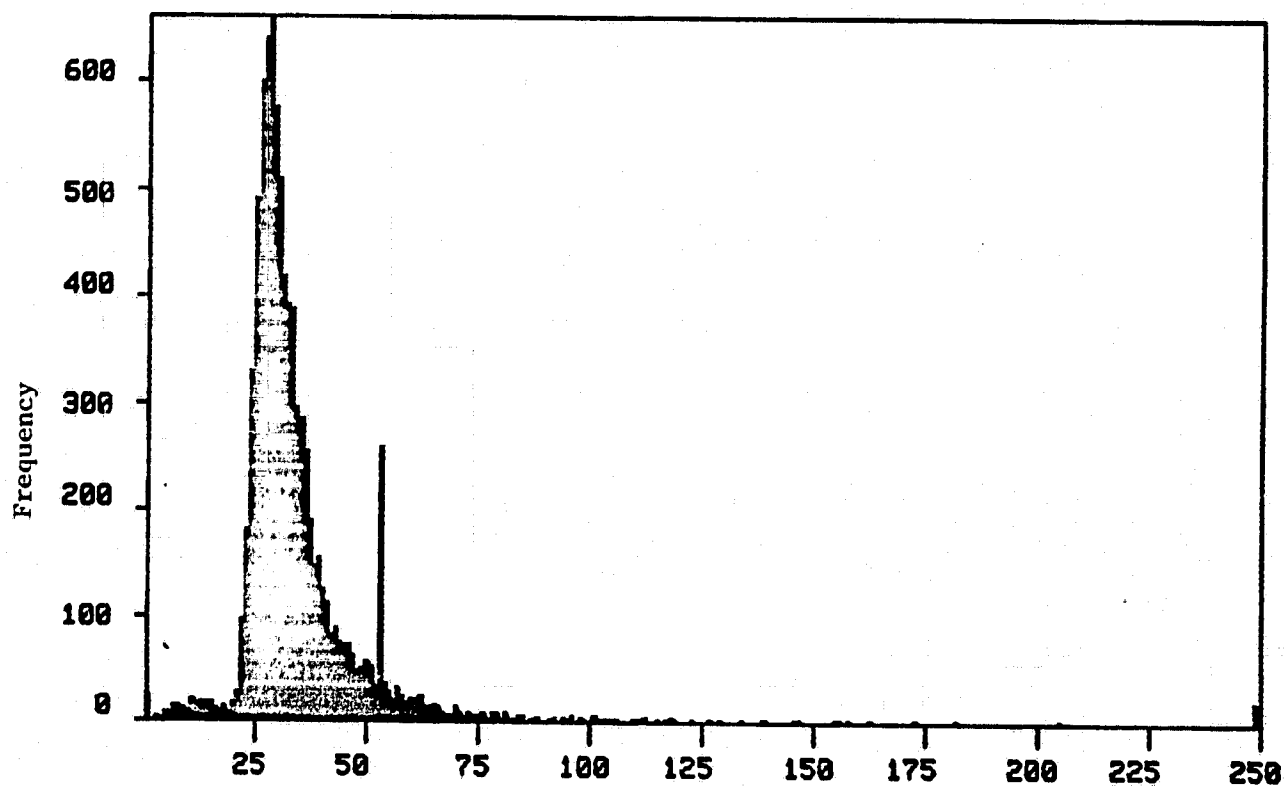


Figure 4. Histograms of  $V$  for Corn and Soybean Component  
Distributions based on Pure Pixels from  
Segment 1380



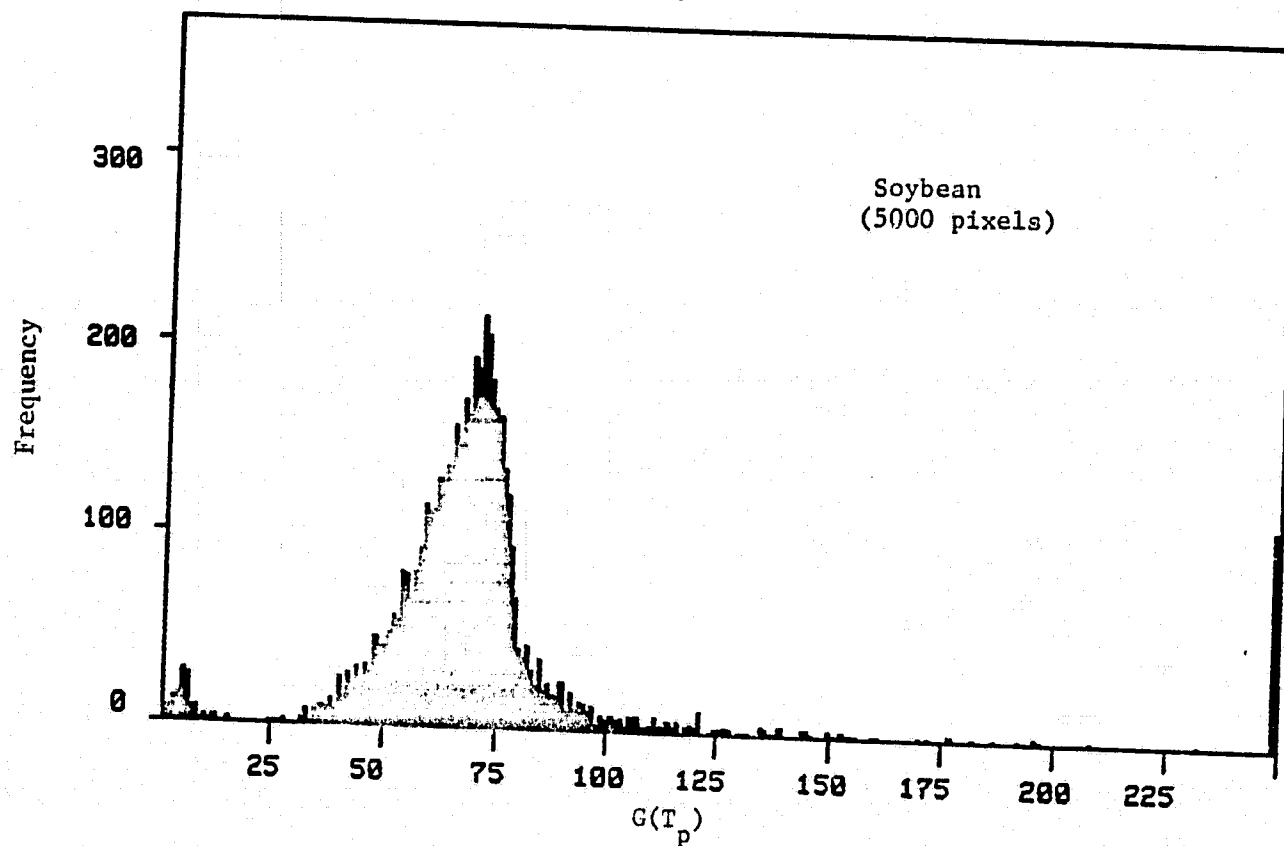
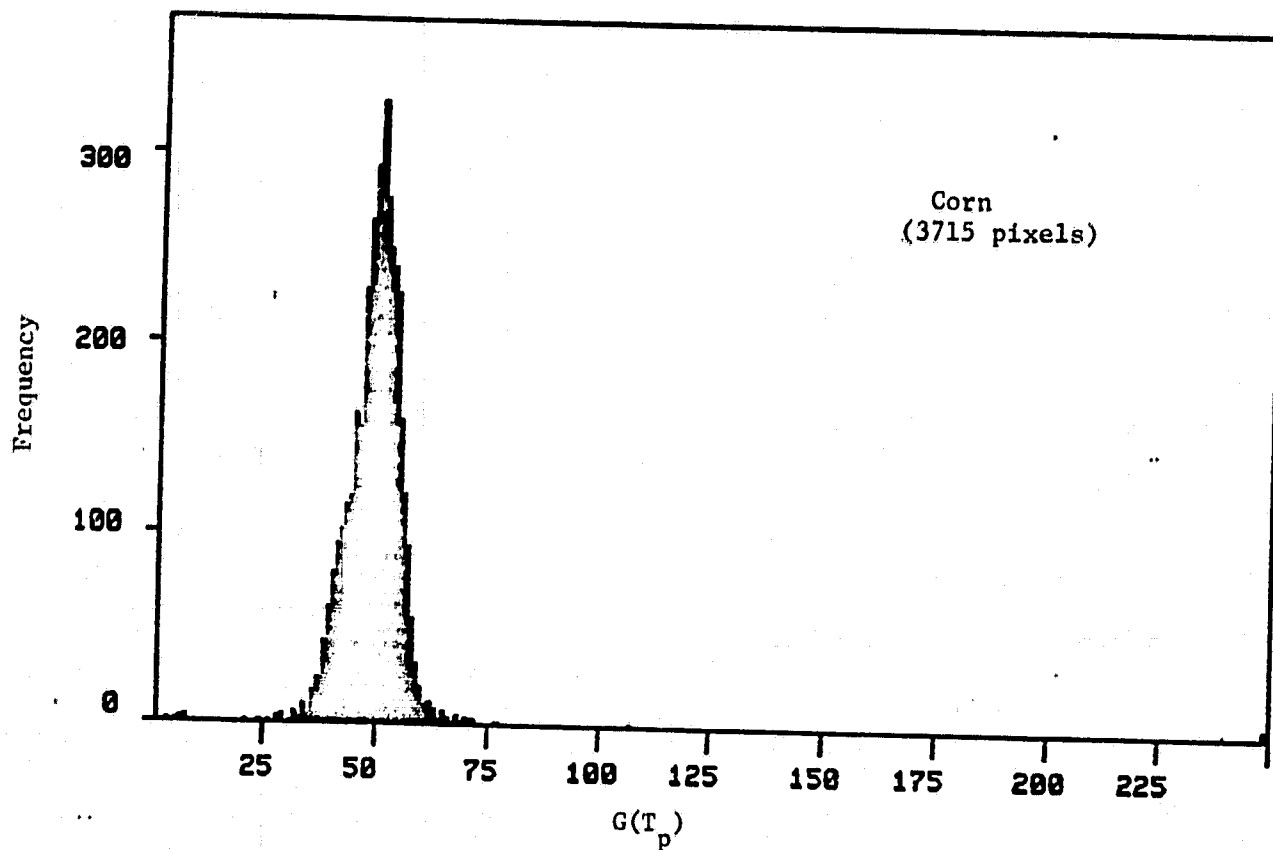
ORIGINAL PAGE IS  
OF POOR QUALITY

Figure 5. Mixture Histogram based on  $\nabla$  for Corn  
and Soybean Components of Figure 4



$\nabla$

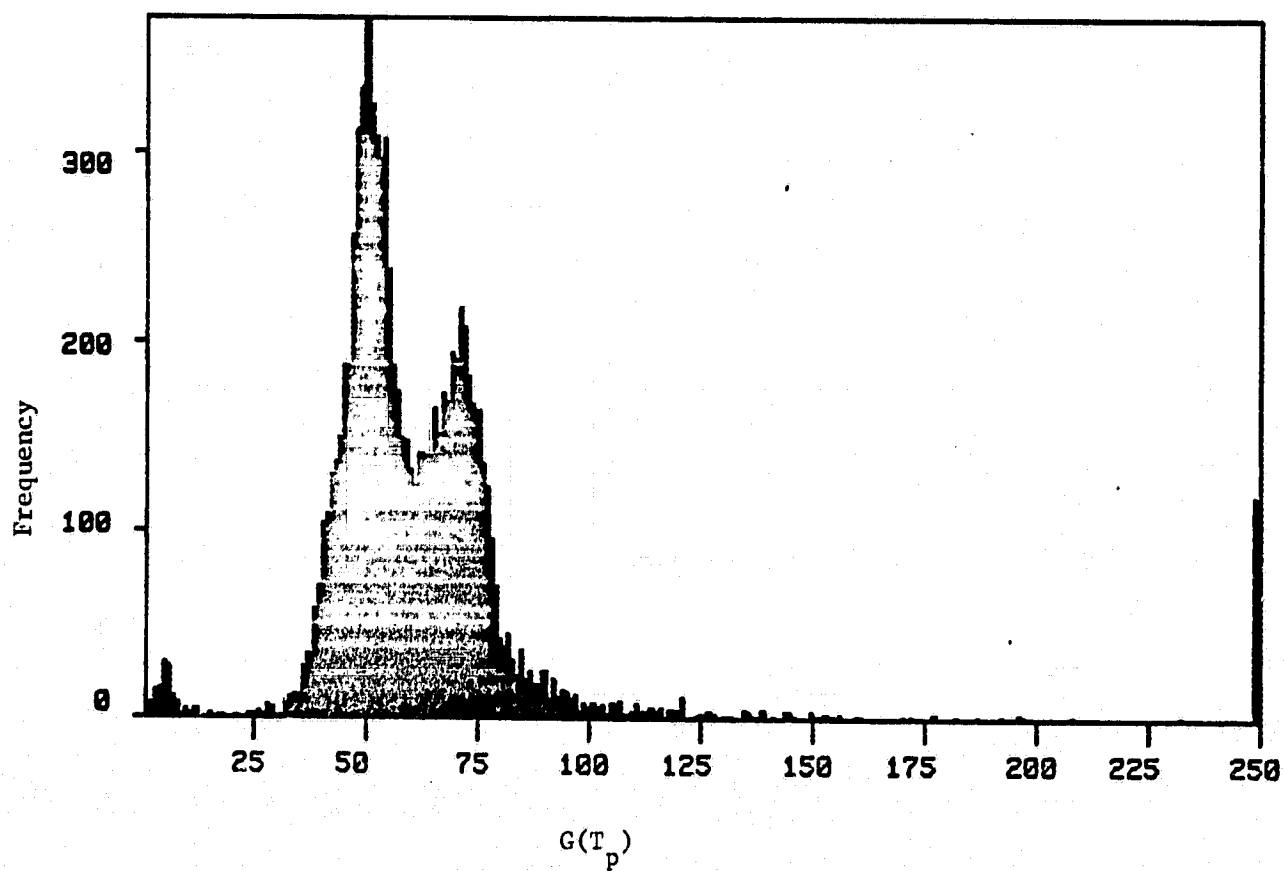
Figure 6. Histograms of  $G(T_p)$  for Corn and Soybean Component  
Distributions based on Pure Pixels from Segment 1380





ORIGINAL PAGE IS  
OF POOR QUALITY

Figure 7. Mixture Histograms based on  $G(T_p)$  for  
Corn and Soybean Components of Figure 6



component distribution for soybeans (see Figure 4). This leads to a second observation which is that the quality of the data is very questionable. The peak in the soybean component should be explained. Further, the figures indicate that outliers are a major problem. For example, note should be made of the extreme values for each profile, particularly for soybean components. In order to correctly analyze these data, the outliers must be more fully understood. Outliers could arise from several sources. Among these are incorrectly specified ground truth readings, crops which were plowed under after the ground truth readings were made, and extreme values which result from instability of parameter estimation in the Badhwar model. Our examination of all of the histograms reveals that outliers are in general most prevalent for  $\nabla$ . We do not at present understand the outliers observed here, but their magnitude is significant to warrant further investigation.

Although the mixtures displayed in Figures 3 and 7 are bimodal, a general impression after examining all of the histograms is that for many of the crop comparisons, the mixture histograms are not bimodal for any of the profile variables. This, of course, causes the usefulness of the profile variables for separating crops to be questioned. Based upon our examination of the data, we are able to make some very general comments concerning the crop separation. For the segments we observed, none of the three variables

produced histograms from which a separation was visible when comparing:

- grass vs. spring small grains
- spring wheat vs. other spring small grains
- spring wheat vs. spring barley
- corn vs. trees
- grass vs. pasture

In contrast, visual separation was present for the following comparisons:

- corn vs. soybeans ( $T_p$  and  $G(T_p)$ )
- cotton vs. spring small grains ( $T_p$  and  $G(T_p)$ )
- sunflower vs. spring wheat ( $T_p$  and  $G(T_p)$ )
- pasture vs. alfalfa ( $G(T_p)$ )

Of course, multivariate examinations of these variables might detect separations which we are unable to observe in the univariate setting.

We also examined the performance of the estimates studied in the simulation studies on the LANDSAT data. In order to do this we sampled from the mixture populations described earlier. Specifically, for selected "crop pair populations" we selected 100 samples of size  $n=200$ , obtaining the MDEN, MLEN, and MDEW for each sample. The results of this "data simulation" were then summarized in much the same way as were the simulations presented earlier. In Table 3 we present the results for estimating the mixing proportion based upon the corn-soybean mixture from Segment

1380. The ground truth proportion there is  $p=.43$  (proportion of pixels in the mixture which are corn). From the table we see that the estimation results are very poor for all estimation procedures. Examination of the histograms (see Figures 2 - 7) reveals that the outliers discussed earlier are probably the major cause for this poor performance. The starting value results deserve special attention. The starting values are restricted to  $p=.1,.2,\dots,.9$ , and they are selected in such a way that if fewer than 5% of the data are "extreme" in either direction, then this has very little effect on the starting values. With outliers as extreme and as numerous as the ones in the present data, the starting value routine often interprets the extreme 10% of the data as constituting a component. Thus we see the extremely poor starting value results in Table 3. In an effort to examine the effect of the outliers on the results in Table 3 we truncated the most extreme observations, and repeated the simulations. In particular, all  $T_p$  observations below 60 and above 150 were truncated, all  $V$  observations above 80 were truncated, and for  $G(T_p)$ , all observations below 10 and above 120 were truncated. These truncations were performed independently for each variable so that the ground truth proportions differ from profile to profile. These ground truth readings are given in Table 3. A truncation based on all three criteria together might be of interest since, for example, the spurious peak at about  $V=53$  for the soybean

Table 3. Results of "Data Simulation" based on Corn and Soybean Pure Pixels from Segment 1380

Sample Size = 200  
Number of Replications = 100

Data Not Truncated  
Ground Truth  $p = .43$  (Proportion Corn)

	$T_p$		$V$		$G(T_p)$	
	$\hat{p}$	$\hat{MSE}$	$\hat{p}$	$\hat{MSE}$	$\hat{p}$	$\hat{MSE}$
MDEN	.59	.07	.64	.05	.63	.10
MLEN	.83	.17	.77	.13	.89	.23
MDEW	.73	.15	.62	.08	.67	.11
Starts	.89	.22	.85	.18	.79	.15

Data Truncated

	Ground Truth $p = .44$		Ground Truth $p = .43$		Ground Truth $p = .45$	
	$\hat{p}$	$\hat{MSE}$	$\hat{p}$	$\hat{MSE}$	$\hat{p}$	$\hat{MSE}$
MDEN	.47	.04	.61	.04	.58	.03
MLEN	.71	.09	.64	.06	.44	.05
MDEW	.49	.09	.57	.05	.54	.02
Starts	.88	.20	.83	.16	.61	.03

data in Figure 4 may be associated with extreme values of one of the other two variables, and thus these values of  $\nabla$  would be removed by such a joint truncation procedure. The results of the simulations on truncated data are given in Table 3. There it can be seen that the performance of the estimators improves dramatically. However, it should be noted that extreme observations seem to continue to have an effect on the results. Notice that  $G(T_p)$  appears to be the best single variable for separating corn and soybeans, in which case MDEW results are superior. Simulations similar to those reported here were obtained for several crop-pair mixtures. In general, although visible separation sometimes existed between the two components, estimation results were usually very poor because of the outliers.

The symmetry of the component distributions is one of our main interests. However, the outliers tend to diminish our ability to examine skewness. Although many of the component distributions appeared to be nearly symmetric, we have observed skewness to the right in several cases, see for example the component distributions for  $\nabla$  in Figure 4.

Many of the comments which are made here are based on our examination of all of the histograms and "data simulation" results which were obtained from our processing of the segment data. Although these displays and results cannot all be included here, we believe that the ones we have presented are sufficient to provide an understanding of

the data. The histograms and data simulation results have been provided to Dr. Dick Heydorn at Johnson Space Center.

### Summary

The results of our investigations have provided new insight into the role of non-normality and the performance of the MLE presently used for crop proportion estimation. In addition we have examined several alternatives to the normal-based MLE for estimating mixing proportions. We believe, however, that further research is needed in this area. In particular, the extension of the investigations to situations in which more than two components are present would be a natural next step. Further extensions to the multivariate case also seem to be of importance.

The MHDE appears to have some real potential as an estimator due to its efficiency under normality. However, much work is necessary before it can be determined whether or not it is a viable alternative.

The role of symmetry of the component distributions and the performance of the estimation procedures still requires examination. In particular, if the asymmetry can be assumed to be in only one direction (probably to the right) then we believe that the estimation results shown in Table 2 can be improved. The practical importance of the

"non-identifiability" observed in the 3-parameter Weibull is not yet fully understood. In addition, possible new alternatives to the Weibull and normal component models considered to date should be considered.

The simulation results concerning the performance of the simple starting value routine we developed imply that further research into its capabilities is warranted.

Finally, the examination of the estimation procedures on LANDSAT data is only in its initial stages. The problem with outliers and how best to deal with them is a very important question related to the implementation of these techniques on LANDSAT data. Although the MDE procedures examined in our investigations are relatively insensitive to outliers, the magnitude and quantity of outliers present in the data we observed had very deleterious effects on all estimation procedures examined.



**APPENDIX A**

**Minimum Distance Estimation of Mixture Model  
Parameters - Asymptotic Results and  
Simulation Comparisons with Maximum Likelihood**

**(SR-63-04427)**

MINIMUM DISTANCE ESTIMATION OF MIXTURE MODEL PARAMETERS -  
ASYMPTOTIC RESULTS AND SIMULATION COMPARISONS  
WITH MAXIMUM LIKELIHOOD

Wayne A. Woodward, William C. Parr,  
William R. Schucany, and Henry L. Gray

1. Introduction

An important problem in aerospace remote sensing is the estimation of the mixing proportions  $p_1, p_2, \dots, p_m$  in the mixture density

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + \dots + p_m f_m(x)$$

where  $m$  is the number of components(crops) in the mixture and for component  $i, f_i(x)$  is a density. The variable of interest,  $X$ , is some measurement such as the reflected energy in four bands of the light spectrum as measured by the LANDSAT satellite, certain linear combinations of these readings, or other derived "feature" variables.

Generally, parameter estimation in mixture model applications has been accomplished by assuming that the component distributions are normal and using maximum likelihood(ML) techniques. In a recent report, Woodward, et. al.(1982) have examined the use of minimum distance(MD) estimation based on the Cramér-von Mises distance, as an alternative to maximum likelihood. Both ML and MD estimation

schemes in that paper were based upon the mixture of two univariate normal distributions whose density function is given by

$$f(x) = \frac{p}{\sqrt{2\pi} \sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + \frac{(1-p)}{\sqrt{2\pi} \sigma_2} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2},$$

where all 5 parameters  $\mu_1$ ,  $\sigma_1$ ,  $\mu_2$ ,  $\sigma_2$ , and  $p$  are unknown. It was also assumed that no training data are available, i.e., the only observations are from the mixture distribution. In this setting, motivated by the crop example,  $p$  is the parameter of paramount importance while location and scale of the components are nuisance parameters. Woodward, et. al. (1982) compare ML and MD estimation techniques on simulated mixtures of normal,  $t(4)$ , and chi-square(9) densities with varying amounts of separation. The results indicate that the MDE is more robust than the MLE to symmetric departures from component normality, while neither technique provides satisfactory results when component distributions are skewed.

In this report, we present further simulation results comparing ML and MD estimation of the mixing proportion based on a mixture-of-normals model, when in fact the component distributions are not normal, yet represent symmetric departures from normality. Unless otherwise indicated, reference to the MDE in this report will involve the use of Cramér-von Mises distance. We also present asymptotic results which establish the strong consistency

and asymptotic normality of MD estimators of the parameters in the mixture-of-normals model, and finally provide asymptotic relative efficiencies for comparing the MLE and MDE in this setting.

## 2. Simulation Results

In this section we report the results of a Monte Carlo study designed to compare the ML and MD estimators based upon a mixture-of-normals when the simulated component distributions are normal and when they are non-normal. These comparisons are made under varying degrees of separation between the two component distributions. All computations were performed on the CDC 6600 at Southern Methodist University.

In these simulations, the mixing proportion,  $p$ , takes on the values .25, .50, and .75. For a given mixture, the component distributions differ from each other only in location and scale. In particular,  $f_1(x)$  is taken to be the density associated with a random variable  $X=aY$  while  $f_2(x)$  is the density for  $X=Y+b$  where  $a>0$ ,  $b>0$ . Thus,  $a$  is the ratio of scale parameters for the densities  $f_1$  and  $f_2$ , and similarly,  $b$  is the difference in location parameters. The random variable  $Y$  in our simulations is either normal, Student's  $t$  with 2 or 4 degrees of freedom, or double exponential. In our simulations we use  $a=1$  and  $a=\sqrt{2}$  while  $b$

is selected to provide the desired separation between the component distributions. The number of modes of the mixture density depends to a large extent on this separation between the two component distributions. Although, for sufficient separation, the mixture model has a characteristic bimodal shape, the density may be unimodal when there is only moderate separation between the components, and in this case, parameter estimation is more difficult than it is in the bimodal cases. For purposes of quantifying this separation between the components, a measure of "overlap" between two distributions was defined by Woodward et. al. (1982).

For each set of parameter configurations, 500 samples of size  $n=100$  were generated from the corresponding mixture distribution. Simulations were based on the IMSL multiplicative congruential uniform random number generator GGUBS. Normal component observations were generated using IMSL subroutine GGNPM which uses the polar method, while  $t(n)$  observations were based on the ratio of independent chi-square and normal deviates, each obtained using IMSL routines. Double exponential components were based on  $\ln(U)$  where  $U$  is uniform(0,1), and randomly assigning either a positive or negative sign. In all cases, observations from the basic component distribution under investigation were simulated and then assigned to either component 1 or component 2 depending upon whether an independent

uniform(0,1) was less than or greater than  $p$ . The observations were then scaled and shifted (with  $a$  and  $b$ ) to provide observations from the appropriate component.

For each sample simulated, both the MDE and MLE were obtained. The iterative procedures discussed by Woodward et. al. (1982) were implemented in such a way that acceptable parameter estimates are obtained for each sample. For example, if the iterative procedure fails to converge in the specified number of iterations, the last value obtained in the iteration is taken to be the estimate if this value is "reasonable" according to preset criteria. In general, if any of the following conditions existed at any step in the iteration,

$$\begin{aligned}\hat{\sigma}_1 &> Y_n - Y_1 \quad (= \text{sample range}) \\ \hat{\sigma}_2 &> Y_n - Y_1 \\ \hat{\mu}_1 &< Y_1 - \frac{Y_n - Y_1}{10} \\ \hat{\mu}_2 &> Y_n + \frac{Y_n - Y_1}{10}\end{aligned}$$

iteration is terminated and the corresponding estimate is taken to be the starting value. This did not occur in any of the 500 repetitions, for most configurations, but did occur a maximum of 7 times out of 500 for MD estimates of the parameters of a mixture of  $t(2)$  components. The extreme observations which occasionally appear in samples from  $t(2)$  mixtures, also forced a modification in the first step of the MLE iteration to avoid a division by zero. Although both

estimation procedures provide estimates of all 5 of the parameters, only the results for estimation of  $p$  will be tabulated since the mixing proportion is the parameter of primary interest, as previously mentioned. In addition, when dealing with the non-normal mixtures, the remaining parameter estimates often do not have a meaningful interpretation.

In Table 1 we present summary results of the simulations comparing the performance of the MLE and MDE for mixtures of normal components while in Table 2 we display the results for the non-normal components. The results for normal and  $t(4)$  components were previously given in Woodward et. al.(1982). Estimates of the bias and MSE based upon the simulations are given by:

$$\hat{\text{Bias}} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{p}_i - p)$$

and

$$\hat{\text{MSE}} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{p}_i - p)^2$$

where  $n_s$  is the number of samples, and  $\hat{p}_i$  denotes an estimate of  $p$  for the  $i$ th sample. It should be noted that  $n\text{MSE}$  is the quantity actually given in the tables since this facilitates comparison with asymptotic variances in Section 4. Since the MLE and MDE are both asymptotically unbiased (this will be discussed for the MDE in the next section),  $n_s \text{MSE} / \sigma^2$  is approximately  $\chi^2(500)$ . It is easy to

Table 1 - Simulation Results for Mixtures of Normal Components

Sample size = 100

Number of Replications = 500

Overlap = .10

Overlap = .03

Ratio of Scale P Factors(a)			Bias	nMSE	E	MDE Closer	Bias	nMSE	E	MDE Closer
.25	1	MDE	.125	7.80	.55	.38	.026	1.09	.49	.39
		MLE	.052	4.26			.008	.539		
		Start	.084	2.06			.048	.782		
.50	1	MDE	.010	3.86	.83	.41	.001	.420	.91	.46
		MLE	.000	3.21			.000	.382		
		Start	-.005	1.22			.001	.634		
.25	$\sqrt{2}$	MDE	.084	5.30	.42	.32	.027	.956	.51	.38
		MLE	.002	2.25			.006	.489		
		Start	-.004	.894			.014	.510		
.50	$\sqrt{2}$	MDE	.005	2.79	.86	.43	.008	.441	.94	.45
		MLE	-.009	2.41			.009	.416		
		Start	-.089	1.85			-.048	.866		
.75	$\sqrt{2}$	MDE	-.137	8.36	.58	.36	-.024	1.08	.44	.42
		MLE	-.086	4.87			-.002	.470		
		Start	-.158	3.97			-.093	1.56		



Table 2. Simulation Results for Mixtures  
of Non-normal Components

Sample size = 100

Number of replications = 500

Double Exponential Components

Overlap = .10

Overlap = .03

Ratio of Scale Factors(a)			Bias	nMSE	E	MDE Closier	Bias	nMSE	E	MDE Closier
.25	1	MDE	.054	2.96	2.13	.66	.030	.545	1.18	.50
		MLE	.091	6.31			.026	.645		
		Start	.065	1.40			.078	1.04		
.50	1	MDE	.007	1.03	4.04	.69	-.001	.286	1.29	.54
		MLE	.007	4.16			-.001	.368		
		Start	-.004	1.17			.000	.414		
.25	$\sqrt{2}$	MDE	.102	4.42	1.40	.60	.035	.775	1.07	.48
		MLE	.034	6.17			.037	.832		
		Start	.011	.926			.050	.678		
.50	$\sqrt{2}$	MDE	.032	1.50	2.71	.68	.003	.259	1.44	.58
		MLE	.073	4.06			.009	.372		
		Start	-.088	1.86			-.035	.570		
.75	$\sqrt{2}$	MDE	-.037	2.20	2.94	.73	-.026	.344	.94	.44
		MLE	-.067	6.47			-.014	.323		
		Start	-.151	3.31			-.107	1.63		

t(4) Components

.25	1	MDE	.104	6.18	1.19	.61	.020	.466	1.89	.49
		MLE	.096	7.35			.029	.883		
		Start	.068	1.59			.072	.998		
.50	1	MDE	.004	1.82	3.07	.69	.000	.266	1.64	.53
		MLE	.015	5.59			-.005	.436		
		Start	.006	1.21			-.001	.496		
.25	$\sqrt{2}$	MDE	.098	5.20	.89	.53	.029	.605	1.61	.49
		MLE	.061	4.63			.044	.976		
		Start	-.010	.810			.036	.654		
.50	$\sqrt{2}$	MDE	.022	1.80	2.77	.67	.001	.300	1.85	.55
		MLE	.028	4.99			.010	.554		
		Start	-.072	1.52			-.046	.778		
.75	$\sqrt{2}$	MDE	-.058	3.68	2.13	.65	-.016	.361	1.57	.50
		MLE	-.076	7.84			-.012	.567		
		Start	-.137	3.07			-.108	1.75		

Table 2 -- Continued

ORIGINAL PAGE IS  
OF POOR QUALITY

t(2) Components

Overlap = .10

Overlap = .03

Ratio of Scale Factors(a)			Overlap = .10				Overlap = .03			
p			Bias	nMSE	E	MDE Closer	Bias	nMSE	E	MDE Closer
.25	1	MDE	.076	3.42	4.30	.80	.024	.308	10.32	.65
		MLE	.199	14.7			.083	3.18		
		Start	.067	1.85			.096	1.37		
.50	1	MDE	-.001	1.34	9.03	.92	-.005	.264	9.24	.63
		MLE	.024	12.1			-.009	2.44		
		Start	-.004	1.39			-.002	.364		
.25	$\sqrt{2}$	MDE	.118	4.92	2.26	.69	.031	.452	7.70	.69
		MLE	.169	11.1			.106	3.48		
		Start	.006	1.18			.071	.962		
.50	$\sqrt{2}$	MDE	.016	1.52	7.76	.89	-.001	.243	8.27	.68
		MLE	.028	11.8			.029	2.01		
		Start	-.078	2.08			-.032	.508		
.75	$\sqrt{2}$	MDE	-.059	2.99	5.79	.85	-.022	.300	11.40	.63
		MLE	-.186	17.3			-.045	3.42		
		Start	-.137	3.37			-.122	1.96		

show then, that the approximate standard error of a tabled  $n\hat{MSE}$  is  $(.0632)(n\hat{MSE})$ . In addition, we also provide the ratio

$$E = \frac{\hat{MSE}(MLE)}{\hat{MSE}(MDE)}$$

as an empirical relative efficiency measure.

In order to take advantage of the paired nature of our ML and MD estimates, we counted the proportion of samples for which  $\hat{p}_D$  is closer to  $p$  than is  $\hat{p}_L$ , where  $\hat{p}_D$  and  $\hat{p}_L$  denote the MD and ML estimates respectively. We present this proportion in the tables under the heading "MDE Closer". This provides an estimate of  $P\{|\hat{p}_D - p| < |\hat{p}_L - p|\}$ . The standard error of the binomial proportions shown in the tables is no greater than  $\sqrt{\frac{(.5)(.5)}{500}} = .022$ .

Analyzing the results, and as can be seen by inspection, we find that the estimated Bias and MSE associated with the MLE were generally smaller than those for the MDE when the components were actually normally distributed. This relationship between the estimators held for both overlaps. The MLE and MDE were quite similar at  $p=.5$  while for  $p=.25$  and  $p=.75$  the superiority of the MLE is more pronounced.

For the mixtures of non-normal components, the relationship between MDE and MLE is reversed in that the MDE generally has the smaller estimated Bias and MSE, especially

for  $t(2)$  mixtures. The superiority of the MDE is due in part to the heavy tails in these components. The MLE often interpreted an extreme observation as being the only sample value from one of the populations with all remaining observations belonging to the other. Due to the well known singularities associated with a zero variance estimate for a component distribution, Day(1969), we were concerned that the observed behavior of the MLE was due to the fact that the variances were not constrained away from zero. However, simulation results in which equal variances were assumed (which removes the singularity) and also those that used a penalized MLE suggested by Redner(1980) were very similar to those quoted here.

A surprising result which was previously noted by Woodward et. al.(1982) is that the starting values obtained using the procedure outlined in Section 3 produced estimators that were competitive with both the MLE and MDE. For both the normal and non-normal mixtures, the MSEs associated with the starting values were generally lower than those for the MDE and MLE when  $\text{overlap}=.10$ . However, when  $\text{overlap}=.03$ , the starting value estimates were generally poorer than those for the MDE and MLE, except for the  $t(2)$  mixtures for which the MLEs were the poorest.

### 3. Asymptotic Distribution Theory for Minimum Cramér-von Mises Distance Estimation

Asymptotic theory for minimum Cramér-von Mises distance estimators for location parameters can be found in Parr and Schucany(1980), and for the general one parameter case in Parr and de Wet(1981). Bolthausen(1977) gives results for the mutiparameter case, but with conditions which are so strict as to rule out scale parameters for unbounded random variables (see his condition III). The purpose of the results in this section is to extend this previous work to cover multiparameter situations including, among others, the problem of normal mixtures.

Assume that at stage  $n$  we observe real-valued  $X_1, X_2, \dots, X_n$  iid from a distribution with cdf  $G$  and let  $G_n$  denote the usual empirical distribution function. Let  $\mathcal{F} = \{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^k\}$ , the projection model, be a family of continuous distribution functions and assume that  $G \in \mathcal{F}$ , i.e.,  $G = F_{\theta_0}$  for some  $\theta_0 \in \Theta$ . Further, assume that there exists an open set  $A \subset \Theta$  with  $\theta_0 \in A$ . Also consider the following continuity(C) and differentiability(D) conditions:

(C) If  $\theta_n \in \Theta$ ,  $n = 1, 2, \dots$ , then

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} (F_{\theta_n}(x) - F_{\theta_0}(x))^2 dF_{\theta_0}(x) = 0$$

implies  $\lim_{n \rightarrow \infty} \theta_n = \theta_0$ .

(D) There exists a function  $\eta: (0,1) \rightarrow \mathbb{R}^k$  such that

$$\sup_{-\infty < x < \infty} |F_{\theta}(x) - F_{\theta_0}(x) - (\theta - \theta_0)' \eta(F_{\theta_0}(x))| = o(\|\theta - \theta_0\|)$$

as  $\|\theta - \theta_0\| \rightarrow 0$ , where  $\|\cdot\|$  is the usual Euclidean norm on  $\mathbb{R}^k$ , and  $\int_0^1 \eta_i^2(u) du < \infty$  for  $i = 1, 2, \dots, k$  where  $\eta'(u) = (\eta_1(u), \eta_2(u), \dots, \eta_k(u))$ .

#### Notes:

1) Condition C is satisfied if, for instance,  $F_{\theta}(x)$  is continuous in  $\theta$  at  $\theta_0$ , pointwise in  $x$  (use dominated convergence). It can be interpreted as requiring that  $\theta$  "continuously parametrize  $\mathcal{F}$ ".

2) If condition C is not satisfied, then this implies  $\sup_{-\infty < x < \infty} |F_{\theta}(x) - F_{\theta_0}(x)|$  can be arbitrarily small without having  $\theta$  approach  $\theta_0$ . In such a case, the search for any consistent estimator seems hopeless. In particular, in such a situation, any consistent estimating functional must be discontinuous with respect to the sup-norm, and hence highly nonrobust.

3) Condition D is weaker than (implied by) quadratic mean differentiability of  $f_{\theta}$  - the canonical regularity condition for asymptotic normality of the maximum likelihood estimator (see LeCam (1970) and Pollard (1980)).

4) Usually,  $\eta_i(u) = \frac{\partial F_{\theta}(x)}{\partial \theta_i} \Big|_{x=F_{\theta}^{-1}(u)}$  and condition D simply

states the uniform validity of the first order Taylor approximation to  $F_\theta(x)$ . If  $k=1$  and  $\theta$  is a location parameter, a sufficient condition to imply D) is that  $F_\theta$  possess a uniformly continuous density.

Before continuing define the  $k \times k$  symmetric matrices  $A$  and  $B$  by

$$A = \{a_{ij}\}, \quad B = \{b_{ij}\}$$

$$\text{with } a_{ij} = \int_0^1 \eta_i(u) \eta_j(u) du$$

$$\text{and } b_{ij} = \int_0^1 \int_0^1 \{\min(u,v) - uv\} \eta_i(u) \eta_j(v) dudv$$

and assume  $A$  to be of full rank. We can now state and outline the proof of the following strong consistency and asymptotic normality results.

Theorem 1: Let  $\theta_n$  be a minimum distance estimator of  $\theta$  for all  $n=1, 2, \dots$ . Then, if condition C holds,  $\theta_n \rightarrow \theta_0$  with probability one.

Proof: Clearly,  $\int (G_n - F_{\theta_0})^2 dF_{\theta_0} \rightarrow 0$  with probability one, and hence also  $\inf_{\theta \in \Theta} \int (G_n - F_\theta)^2 dF_\theta \rightarrow 0$  with probability one.

Now,

$$\sup_{\theta} \left| \int (G_n - F_\theta)^2 dF_\theta - \int (F - F_\theta)^2 dF_\theta \right| \leq 4 \sup_{-\infty < t < \infty} |G_n(t) - F_{\theta_0}(t)| \rightarrow 0$$

with probability one. Hence,

$$\int (F_{\theta_0} - F_{\theta_n})^2 dF_{\theta_n} = \int (F_{\theta_n} - F_{\theta_0})^2 dF_{\theta_0} \rightarrow 0$$

with probability one, and strong consistency of  $\theta_n$  follows from the assumption.

Theorem 2: Assume conditions C and D and that A is of full rank. Then, if  $f_\theta(x)$  is continuous in  $\theta$  at  $\theta_0$  for every  $x$ ,

$$\sqrt{n} (\theta_n - \theta_0) \xrightarrow{L} N(0, A^{-1}BA^{-1}).$$

Proof. (Sketched)

Set

$$K_n(\xi) = n \int (G_n - F_{\theta_0 + \xi/\sqrt{n}})^2 dF_{\theta_0 + \xi/\sqrt{n}} \quad \text{for } \xi \in R^k.$$

Then we have

$$\begin{aligned} K_n(\xi) &= n \int (G_n - F_{\theta_0} - (F_{\theta_0 + \xi/\sqrt{n}} - F_{\theta_0}))^2 dF_{\theta_0} \\ &\quad + n \int (G_n - F_{\theta_0} - (F_{\theta_0 + \xi/\sqrt{n}} - F_{\theta_0}))^2 d[F_{\theta_0 + \xi/\sqrt{n}} - F_{\theta_0}] \\ &= o_p(1) + \int_0^1 (U_n(t) - \xi' n(t) - R_n(t))^2 dt, \end{aligned}$$

uniformly in  $\xi$  for  $\xi' \xi \leq C$ , for any  $C < \infty$ , where

$\sup_{0 < t < 1} |R_n(t)| \rightarrow 0$  with probability one, also uniformly in  $\xi$  for  $\xi' \xi \leq C$ . Here,  $U_n(t) = \sqrt{n}(G_n(F_{\theta_0}^{-1}(t)) - t)$ ,  $0 < t < 1$ .

By an extension of the argument of Pyke (1970, p. 29-30) to the present context, we obtain that the limiting law of the random variable minimizing  $K_n(\xi)$  over  $\xi$  is also that of the value minimizing

$$\int_0^1 (B(t) - \xi' n(t))^2 dt,$$

where  $B$  is a Brownian bridge. The result then follows immediately.

It can be shown that the mixture of normals model satisfies the conditions of both Theorem 1 and Theorem 2.



#### 4. Asymptotic Relative Efficiencies

Theorem 2 of the previous section indicates that for the mixture-of-normals model, we have

$$\sqrt{n} (\theta_n - \theta_0) \xrightarrow{L} N(\underline{0}, A^{-1}BA^{-1}) ,$$

where  $\theta_0 = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p)$  and  $\theta_n$  is the vector of corresponding MD estimators using Cramér-von Mises distance. Likewise, it is well known that

$$\sqrt{n} (\hat{\theta}_L - \theta_0) \xrightarrow{L} N(\underline{0}, I^{-1}(\theta_0)) ,$$

where  $\hat{\theta}_L$  is the MLE of  $\theta_0$  and  $I(\theta_0)$  is Fisher's information matrix. We will employ the usual terminology and refer to  $A^{-1}BA^{-1}$  and  $I(\theta_0)$  as asymptotic variance - covariance matrices and to their diagonal elements as asymptotic variances of the corresponding estimators. In this section we will present computed asymptotic variances for the MDE of  $p$ , which is denoted by  $\hat{p}_D$ , and compare these with the asymptotic variances associated with the MLE, denoted by  $\hat{p}_L$ .

The components of the matrix  $A$  were evaluated using the expression

$$\int_{-\infty}^{\infty} \xi_i(x) \xi_j(x) f_{\theta}(x) dx ,$$

where  $F_{\theta}(x)$  and  $f_{\theta}(x)$  denote the distribution function and density function respectively for the mixture,  $\theta_i$  is the  $i$ th

component of  $\theta$ , and

$$\xi_i(x) = \frac{\partial F_\theta(x)}{\partial \theta_i}.$$

This integral was evaluated using IMSL subroutine DCADRE which employs Romberg extrapolation to perform numerical integration of an integral over a finite interval. In our implementation, we used DCADRE to evaluate the integral

$$\int_U^L \xi_i(x) \xi_j(x) f_\theta(x) dx,$$

where  $L = \min(-10\sigma_1 + \mu_1, -10\sigma_2 + \mu_2)$  and  $U = \max(10\sigma_1 + \mu_1, 10\sigma_2 + \mu_2)$  with maximum allowable absolute error specified as  $1.0 \times 10^{-15}$  and relative error of  $1.0 \times 10^{-12}$ . The double integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{F_\theta(\min(x,y)) - F_\theta(x)F_\theta(y)\} \xi_i(x) \xi_j(y) f_\theta(x) f_\theta(y) dx dy$$

involved in calculating the elements of the matrix  $B$  is approximated by using IMSL subroutine DBLIN to perform a Romberg integration of the integral

$$\int_L^U \int_L^U \{F_\theta(\min(x,y)) - F_\theta(x)F_\theta(y)\} \xi_i(x) \xi_j(y) f_\theta(x) f_\theta(y) dx dy$$

with maximum allowable absolute error specified as  $1.0 \times 10^{-9}$ .

The calculation of the information matrix for the

mixture-of-normals model is discussed by Behboodian(1972). We have followed Behboodian's procedure and used Gauss-Hermite quadrature to approximate the integrals involved. Using 48-point quadrature we obtain good agreement with Behboodian's tabled results.

In Table 3 we display the asymptotic variances for  $\hat{p}_D$  and  $\hat{p}_L$  along with asymptotic relative efficiency (ARE) calculated as

$$\text{ARE} = \frac{\text{asymptotic variance } \hat{p}_L}{\text{asymptotic variance } \hat{p}_D} .$$

These values are calculated for each of the parameter configurations employed in Table 1 for the normal mixtures. As in Table 1, the asymptotic results indicate that the MDE compares more favorably with the MLE when  $p=.5$  while its relative performance is not as good for  $p=.25$  or  $p=.75$ .

Table 3 - Asymptotic Relative Efficiencies

			Overlap = .10		Overlap = .03	
p	Ratio of Scale Factors(a)		Asymptotic Variance	ARE	Asymptotic Variance	ARE
.25	1	MDE	13.60 (7.80)*	.42 (.55)	.471 (1.09)	.69 (.49)
		MLE	5.67 (4.26)		.323 (.539)	
.50	1	MDE	4.54 (3.86)	.65 (.83)	.398 (.420)	.89 (.91)
		MLE	2.95 (3.21)		.355 (.382)	
.25	$\sqrt{2}$	MDE	18.77 (5.30)	.32 (.42)	.511 (.956)	.65 (.51)
		MLE	5.96 (2.25)		.330 (.489)	
.50	$\sqrt{2}$	MDE	3.49 (2.79)	.68 (.86)	.395 (.441)	.89 (.94)
		MLE	2.39 (2.41)		.353 (.416)	
.75	$\sqrt{2}$	MDE	5.51 (8.36)	.58 (.58)	.420 (1.08)	.73 (.44)
		MLE	3.18 (4.87)		.305 (.470)	

\*Associated Monte Carlo results from Table 1 are given in parentheses.

## 5. Concluding Remarks

We believe that the results of this paper provide further evidence that the use of the MDE should be considered in crop proportion estimation procedures developed by NASA. Our results, again, and more conclusively than before, indicate that the MDE is indeed more robust than the MLE in the sense that it is less sensitive to symmetric departures from the underlying assumption of normality of component distributions.

Woodward et. al. (1983) have investigated basing the MD estimation procedure on a mixture of Weibull components in order to allow for possible asymmetry in the component distributions. Their results indicate that this approach provides a viable alternative to the normal-based procedures discussed here. Research is also proceeding on the case of multiple ( $>2$ ) components in the mixture.

The results of Section 4 indicate that the MDE does not perform as well as would be hoped when the data actually do arise from a mixture-of-normals model. We are currently examining the use of the Hellinger metric in this regard due the results of Beran(1977) concerning the full asymptotic relative efficiency of minimum Hellinger distance estimators.

## REFERENCES

1. Behboodian, J. (1972). "Information Matrix for a Mixture of Two Normal Distributions," J. Statist. Comput. Simul. 1, 295-314.
2. Beran, R. (1977). "Minimum Hellinger Distance Estimates for Parametric Models," Annals of Statistics 5, 445-463.
3. Bolthausen, E. (1977). "Convergence in Distribution of Minimum-Distance Estimators," Metrika, 24, 215-228.
4. Day, N. E. (1969). "Estimating the Components of a Mixture of Normal Distributions," Biometrika, 56, 463-474.
5. LeCam, L. (1970). "On the Assumptions Used to Prove Asymptotic Normality of Maximum Likelihood Estimates," Annals Math. Statist., 41, 802-828.
6. Parr, W.C. and De Wet, T. (1981). "On Minimum Cramer-von Mises-norm Parameter Estimation," Communications in Statistics, A10, 1149-1166.
7. Parr, W. C. and Schucany W. R. (1980). "Minimum Distance and Robust Estimation," J. Amer. Statist. Assoc., 75, 616-624.
8. Pollard, D. (1980). "The Minimum Distance Method of Testing," Metrika, 27, 43-70.

9. Pyke, R. (1970). "Asymptotic Results on Rank Statistics," Nonparametric Techniques in Statistical Inference, ed. by M.L.Puri, Cambridge University Press, London.
10. Redner, R.A. (1980). "Maximum Likelihood Estimation for Mixture Models," JSC-16832, October 1980.
11. Woodward, W.A., Gunst, R.F., Lindsey, H., and Gray, H.L. (1983). "Proportion Estimation in Mixtures of Asymmetric Distributions," NASA Technical Report SR-63-04409, May 1983.
12. Woodward, W.A., Schucany, W.R., Lindsey, H. and Gray, H.L. (1982). "A Comparison of Minimum Distance and Maximum likelihood Techniques for Proportion Estimation," NASA technical report SR-62-04376, November 1982.

**APPENDIX B**

**Minimum Hellinger Distance Estimation  
of Mixture Model Parameters**

**(SR-63-04433)**



# MINIMUM HELLINGER DISTANCE ESTIMATION OF MIXTURE MODEL PARAMETERS

Wayne A. Woodward and Paul W. Eslinger

## 1. Introduction

Recent reports by Woodward et. al. (1982,1983) have considered minimum distance estimation(MCVMDE), based on Cramér-von Mises distance, as an alternative to maximum likelihood(ML) for estimating the parameters of the mixture-of-normals model. Their results indicate that the MCVMDE is more robust to departures from the assumption of normal components than is maximum likelihood. In particular, they have shown that if mixture-of-normal based MCVMD and ML procedures are used to estimate the parameters of a mixture of symmetric (but non-normal) distributions such as double exponential,  $t(4)$ , or  $t(2)$ , then the MCVMDE produces superior proportion estimates. However, their results also show that when the component distributions actually are normal, the MLE is superior.

Intuitively, robust procedures are those which are insensitive to small deviations from the assumptions. Typically, robust procedures obtain this robustness at the expense of not being optimal at the true model. In fact, Bickel(1978) describes robustness as "paying a price in

terms of efficiency at the (true) model in terms of reasonably good maximum M.S.E. over the neighborhood." The behavior of the MCVME described above is a good example of this trade-off. However, Beran(1977) has suggested the use of the minimum Hellinger distance(MHD) estimator which has certain robustness properties and is asymptotically efficient at the true model. Its applicability to aerospace remote sensing is of interest since it has the potential of providing robust proportion estimates under deviations from normality while maintaining performance comparable to the MLE when the underlying components actually are normal. In this report we will briefly examine the use of the MHDE for estimating the parameters of the mixture-of-normals model.

## 2. The Minimum Hellinger Distance Estimator

Let  $X_1, X_2, \dots, X_n$  denote a random sample from some unknown distribution and let  $Y_1, Y_2, \dots, Y_n$  denote the corresponding order statistics. Further, Let  $\mathcal{F}_\theta = \{F_\theta: \theta \in \Theta\}$  be a family of distributions, called the projection family or projection model, depending on the (possibly vector valued) parameter  $\theta$ . A minimum distance estimate of  $\theta$  is a value  $\hat{\theta}$  which minimizes the distance between the data distribution (whose model is unknown) and the projection model. In particular, the MCVME minimizes the Cramér-von

Mises distance between the empirical distribution function and  $F_\theta$ . For more discussion, the reader is referred to Woodward, et.al.(1982).

Hellinger distance between two absolutely continuous distributions is defined to be  $||f^{1/2}-g^{1/2}||$  where  $f$  and  $g$  are the corresponding densities and  $||.||$  denotes the usual  $L^2$  norm, i.e.

$$||f^{1/2}-g^{1/2}|| = [\int (f^{1/2}-g^{1/2})^2 dx]^{1/2} \quad (2.1)$$

where integration is with respect to Lebesgue measure on the real line. Let  $\mathcal{F}$  denote the set of all absolutely continuous probability functions with respect to Lebesgue measure on the real line, and for our purposes, let  $\mathcal{F}_\theta = \{F_\theta: \theta \in \Theta\}$ , the projection family, be a parametrized subset of  $\mathcal{F}$ . The MHD estimator  $\hat{\theta}_H$  of  $\theta$  is defined as a value of  $\theta$  which minimizes  $||f_\theta^{1/2}-\hat{g}_n^{1/2}||$  where  $\hat{g}_n$  is a suitable nonparametric density estimator. It should be noted that minimizing  $||f_\theta^{1/2}-\hat{g}_n^{1/2}||$  is equivalent to maximizing

$$\int f_\theta^{1/2} \hat{g}_n^{1/2} dx \quad (2.2)$$

and we will utilize this form for computational convenience.

Beran(1977) and Stather(1982) have provided theoretical results establishing the consistency, asymptotic normality, asymptotic full efficiency, and robustness of the MHDE. However, their results only briefly discuss the computational aspects of implementing the MHDE and provide

only limited empirical evidence concerning its robustness. In this report we investigate the usefulness of the MHDE for estimating the parameters of the mixture-of-normals model.

In the mixture-of-normals setting,  $f_{\theta}$  becomes

$$f_{\theta}(x) = \frac{p}{\sqrt{2\pi} \sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + \frac{(1-p)}{\sqrt{2\pi} \sigma_2} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} \quad (2.3)$$

where  $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, p)'$ . In the next section we present the results of a simulation study in which the MHDE is calculated using the projection model in (2.3). In these calculations, we have employed Newton's method to maximize (2.2), which produces the iterative algorithm

$$\hat{\theta}_H^{(m+1)} = \hat{\theta}_H^{(m)} - \left[ \int \frac{\partial^2 f_{\theta}^{1/2}}{\partial \theta^2} \hat{g}_n^{1/2} dx \right]^{-1} \int \frac{\partial f_{\theta}^{1/2}}{\partial \theta} \hat{g}_n^{1/2} dx \quad (2.4)$$

where  $\hat{\theta}_H^{(m)}$  denotes the estimate of  $\theta$  obtained on the  $m$ th step, and  $\hat{\theta}_H^{(0)}$  denotes the starting value,  $(\hat{\mu}_1^{(0)}, \hat{\sigma}_1^{(0)}, \hat{\mu}_2^{(0)}, \hat{\sigma}_2^{(0)}, \hat{p}^{(0)})$ .

If any step produces estimates of  $\sigma_1$  or  $\sigma_2$  which are less than zero, then we use a scaled step "half-way" to zero.

In the implementation the density estimator used is

$$\hat{g}_n^{1/2}(x) = \frac{1}{nc_n s_n} \sum_{i=1}^n w\left(\frac{x-X_i}{c_n s_n}\right)$$

based on the Epanechnikov kernel  $w(x) = .75(1-x^2)$  for  $|x| \leq 1$ , with the scale statistic  $s_n$  set to  $\hat{\sigma}_1^{(0)}$  when  $\hat{p}^{(0)} \geq .5$  and  $\hat{\sigma}_2^{(0)}$  when  $\hat{p}^{(0)} < .5$ . For a discussion of density estimators see Tapia and Thompson (1978). The value for  $c_n$  is given by the expression  $c_n = 2.16n^{-.271}$ . These values of  $c_n$  are optimal for use with a normal projection model and are used here for

convenience. Although further investigation into  $c_n$  values which are optimal for use with the projection model of (2.3) is needed, we believe that the  $c_n$  values utilized are sufficient for the purpose here. When the projection model in (2.3) is used, it follows that  $\int \frac{\partial f_{\theta}^{1/2}}{\partial \theta} \hat{g}_n^{1/2} dx$  in (2.4) is a  $5 \times 1$  vector while  $\int \frac{\partial^2 f_{\theta}^{1/2}}{\partial \theta^2} \hat{g}_n^{1/2} dx$  in (2.4) is a  $5 \times 5$  matrix, the elements of which are integrals to be evaluated at each step of the iterative procedure. In the Appendix we show the partial derivatives involved in the calculation of  $\frac{\partial f_{\theta}^{1/2}}{\partial \theta}$  and  $\frac{\partial^2 f_{\theta}^{1/2}}{\partial \theta^2}$ . We have chosen to evaluate the numerical integrals using the trapezoidal rule over a grid of 100 steps equally spaced between  $Y_1 - c_n s_n$  and  $Y_n + c_n s_n$ , i.e. the range of support of  $\hat{g}_n$ .

### 3. Simulation Results

In this section we report the results of simulations designed to provide empirical evidence concerning the effectiveness of the MHDE using a mixture-of-normals projection model when the component distributions in the simulated samples are normal and when they are non-normal. In addition, we have made our comparisons under two levels of separation between the component distributions.

In these simulations, we have used parameter configurations previously considered by Woodward, et. al.

(1983). In particular, we use mixing proportions .25, .50, and .75 and "overlaps" as defined by Woodward, et. al.(1982) of .03 and .10. Again, as in the previous work, we consider cases in which the ratio of the standard deviations of component 1 to component 2 is 1 and when it is  $\sqrt{2}$ . In these simulations we have simulated mixtures with normal and  $t(4)$  components. For each set of configurations, 500 samples of size  $n=100$  were generated from the corresponding mixture distribution. Simulations were performed on the CDC 760 computer. Starting values were obtained as discussed by Woodward, et. al.(1982) with the exception that starting values for the component standard deviations,  $\sigma_1$  and  $\sigma_2$ , utilized in this study are smaller than those used in the previous reports (Woodward, et. al.(1982,1983)) by a factor of approximately 1.2. For each sample simulated, the MCVME, MHDE, and MLE for all 5 parameters were obtained. However, only the results for the estimation of  $p$  are tabled since the mixing proportion is the parameter of interest.

In Table 1 we present the results for simulated mixtures of normal components, while in Table 2 we show the results for simulated mixtures of  $t(4)$  components. Simulation based estimates of the bias and MSE associated with the various estimators are given by

$$\hat{\text{Bias}} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{p}_i - p)$$

and

$$\hat{\text{MSE}} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{p}_i - p)^2$$

Table 1 - Simulation Results for Mixtures of Normal Components

Sample size = 100  
Number of Replications = 500

Overlap = .03

Overlap = .10

P	Ratio of Scale Factors(a)		$\hat{\text{Bias}}$	$\hat{\text{nMSE}}$	$\hat{\text{E}}$	$\hat{\text{Bias}}$	$\hat{\text{nMSE}}$	$\hat{\text{E}}$
.25	1	MCVMDE	.1271	8.060	.594	.0286	1.050	.420
		MHDE	.0832	3.848(385)	1.245	.0184	.522(418)	.854
		MLE	.0611	4.790		.0067	.446	
		Start	.0728	1.942		.0516	.826	
.50	1	MCVMDE	-.0002	3.717	.856	.0019	.444	.995
		MHDE	-.0043	2.407(449)	1.321	.0034	.487(473)	.908
		MLE	.0033	3.180		.0021	.442	
		Start	.0004	1.096		-.0012	.736	
.25	$\sqrt{2}$	MCVMDE	.0950	5.891	.408	.0371	1.219	.365
		MHDE	.0013	1.483(383)	1.622	.0092	.390(437)	1.167
		MLE	.0043	2.405		.0108	.455	
		Start	-.0124	.914		.0090	.510	
.50	$\sqrt{2}$	MCVMDE	.0065	3.066	.794	.0022	.468	.857
		MHDE	-.0333	1.963(460)	1.239	-.0112	.556(480)	.721
		MLE	-.0196	2.433		.0011	.401	
		Start	-.0894	1.778		-.0560	.996	
.75	$\sqrt{2}$	MCVMDE	-.1423	8.571	.624	-.0288	1.194	.378
		MHDE	-.1199	4.711(405)	1.135	-.0392	.978(410)	.461
		MLE	-.0897	5.348		-.0085	.451	
		Start	-.1806	4.346		-.0936	1.530	

Table 2 - Simulation Results for Mixtures of  $t(4)$  Components

Sample size = 100  
Number of Replications = 500

Overlap = .03

Overlap .10

P	Ratio of Scale Factors(a)	Overlap .10					Overlap = .03				
		$\hat{\text{Bias}}$	$\hat{\text{nMSE}}$	$\hat{\text{E}}$	$\hat{\text{Bias}}$	$\hat{\text{nMSE}}$	$\hat{\text{Bias}}$	$\hat{\text{nMSE}}$	$\hat{\text{E}}$	$\hat{\text{Bias}}$	$\hat{\text{E}}$
.25	MCVMDE	.0811	4.755	1.405	.0206	.524	.0206	.524	1.410	.0206	1.410
	MHDE	.0508	2.476(395)	2.698	.0286	.580(414)	.0286	.580(414)	1.274	.0286	1.274
	MLE	.0912	6.680		.0241	.739	.0241	.739		.0241	
	Start	.0736	1.638		.0670	.962	.0670	.962		.0670	
.50	MCVMDE	-.0024	1.959	2.895	.0037	.308	.0037	.308	1.594	.0037	1.594
	MHDE	.0015	1.746(437)	3.249	.0039	.333(488)	.0039	.333(488)	1.474	.0039	1.474
	MLE	.0059	5.672		.0032	.491	.0032	.491		.0032	
	Start	.0036	1.064		.0006	.490	.0006	.490		.0006	
.25	MCVMDE	.0954	4.868	1.060	.0309	.727	.0309	.727	1.479	.0309	1.479
	MHDE	.0235	2.038(384)	2.532	.0178	.456(411)	.0178	.456(411)	2.357	.0178	2.357
	MLE	.0676	5.161		.0481	1.075	.0481	1.075		.0481	
	Start	.0024	.902		.0398	.618	.0398	.618		.0398	
.50	MCVMDE	.0159	1.878	3.127	-.0019	.301	-.0019	.301	1.588	-.0019	1.588
	MHDE	-.0490	2.316(423)	2.536	-.0077	.320(485)	-.0077	.320(485)	1.494	-.0077	1.494
	MLE	-.0128	5.873		.0076	.478	.0076	.478		.0076	
	Start	-.0874	1.890		-.0422	.662	-.0422	.662		-.0422	
.75	MCVMDE	-.0483	3.404	2.401	-.0161	.409	-.0161	.409	1.782	-.0161	1.782
	MHDE	-.0985	4.241(384)	1.927	-.0487	1.053(391)	-.0487	1.053(391)	.692	-.0487	.692
	MLE	-.0742	8.172		-.0119	.729	-.0119	.729		-.0119	
	Start	-.1286	2.614		-.1012	1.558	-.1012	1.558		-.1012	



where  $n_s$  denotes the number of samples and  $\hat{p}_i$  denotes an estimate of  $p$  for the  $i$ th sample. As in the earlier reports,  $n\hat{MSE}$  is given in the table where  $n$  is the size of each individual sample (in our case 100). We provide the ratios

$$\hat{E}_{CVM} = \frac{\hat{MSE}(MLE)}{\hat{MSE}(MCVMDE)}$$

and

$$\hat{E}_H = \frac{\hat{MSE}(MLE)}{\hat{MSE}(MHDE)}$$

as empirical measures of the relative efficiencies of the MCVMDE and MHDE respectively with the MLE. An approximate standard error of a tabled  $n\hat{MSE}$  is  $(.0632)(n\hat{MSE})$ .

The results in Tables 1 and 2 illustrate the characteristics of the MHDE shown theoretically by Beran(1977) and Stather(1982). In particular, for the simulated mixtures of normal components in Table 1, the MSEs for the MHDE were comparable (in most instances) to those for the MLE and smaller than those for the MCVMDE. This behavior can also be seen by noting that  $\hat{E}_H$  is close to 1 for most configurations while  $\hat{E}_{CVM}$  is consistently less than 1. However, in Table 2, for simulated mixtures of  $t(4)$  components,  $\hat{E}_H$  was greater than 1 in all but one case. In addition, the robustness shown by the MHDE was in most cases comparable to that for the MCVMDE as evidenced by similar values of  $\hat{E}_H$  and  $\hat{E}_{CVM}$ . As noted in the previous reports, see Woodward, et. al.(1982,1983), the starting value routine provided good estimates, which in fact were competitive with those given by ML, MCVM, and MHD techniques.

A few further comments are in order. First, although the computational aspects of the MHDE are complex, we found the computer time required for the MHDE to be similar to that for the other two estimators. The Newton-Raphson procedure used to calculate the MHDE is quadratically convergent. This usually resulted in convergence within 10 steps for the MHDE. The MCVME also usually converged within 10 iterations while the MLE required more, especially for the .10 overlap, in which case more than 50 steps were often required. However, the MLE is computationally much simpler at each step. For a discussion of the computational procedures used to calculate the MLE and MCVME, see Woodward et. al. (1982,1983).

The number in parentheses after the MHDE results in the table is the number of times (out of 500) that the MHDE actually converged. When convergence was not obtained for any of the estimators, the estimate was taken to be the starting value. For the MCVME and MLE, convergence was almost always obtained. However, it can be seen that the failure of the MHDE to converge was a common occurrence. Of course, the results in the tables for the MHD must be viewed accordingly, i.e. approximately 20% of the "MHD" estimates used in the bias and MSE calculations are actually starting values. In some instances, this may improve the performance of the MHDE.

A related observation is that the MHDE seems to be

quite sensitive to starting values. For example, in the tables, we see that the poorest results for the MHDE are obtained when  $p=.75$  and the ratio of standard deviations between components is  $\sqrt{2}$ . It should be noted that this is also the situation in which the starting values are the poorest. While the other two estimators do not seem to be overly affected by these poor starts, the MHD is quite sensitive. As noted earlier, the starting values for  $\sigma_1$  and  $\sigma_2$  used here are smaller than the intuitively appealing ones proposed earlier by Woodward, et. al., (1982,1983). Although we do not understand why, the use of these smaller starting values improves the performance of the MHDE (and has very little effect on the MLE and MCVME).

In related investigations of the MHDE, we have examined its performance on the estimation of the location and scale parameters of a univariate normal projection model. In this setting we have also seen an extreme sensitivity to starting values. In Table 3a we display an array of starting values for  $\mu$  and  $\sigma$  of a univariate normal projection model. Samples of size  $n=40$  were simulated from a normal distribution with  $\mu=0$  and  $\sigma=1$ . In Table 3b we provide an associated array displaying the number of times out of 1000 such samples that the iterative routines for the MHDE converged when using the corresponding starting values in the array of Table 3a. The sensitivity of the MHDE to poor starting values is very evident. It should be noted that using the "good" starting

Table 3 - Effect of Starting Value on  
MHD Estimators  
of  $\mu$  and  $\sigma$  from 1000 Simulated  
N(0,1) Samples of Size  $n = 40$

(a)  
Starting Values  
 $(\hat{\mu}^{(0)}, \hat{\sigma}^{(0)})$

$(-1, \frac{1}{2})$	$(-1, \sqrt{\frac{1}{2}})$	$(-1, 1)$	$(-1, \sqrt{2})$	$(-1, 2)$
$(-\frac{1}{2}, \frac{1}{2})$	$(-\frac{1}{2}, \sqrt{\frac{1}{2}})$	$(-\frac{1}{2}, 1)$	$(-\frac{1}{2}, \sqrt{2})$	$(-\frac{1}{2}, 2)$
$(0, \frac{1}{2})$	$(0, \sqrt{\frac{1}{2}})$	$(0, 1)$	$(0, \sqrt{2})$	$(0, 2)$
$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, \sqrt{\frac{1}{2}})$	$(\frac{1}{2}, 1)$	$(\frac{1}{2}, \sqrt{2})$	$(\frac{1}{2}, 2)$
$(1, \frac{1}{2})$	$(1, \sqrt{\frac{1}{2}})$	$(1, 1)$	$(1, \sqrt{2})$	$(1, 2)$

(b)

Number of Times (out of 1000) that  
MHDE Converged Using Starting  
Values from Table 3a

62	43	59	458	284
210	420	867	866	233
834	993	999	876	179
196	415	843	866	224
71	49	60	463	265

values  $\hat{\mu}^{(0)} = \text{median}$  and  $\hat{\sigma}^{(0)} = \text{median}\{|X_i - \hat{\mu}^{(0)}|\} / .6745$ , obtained from the data for each sample, resulted in convergence of the MHDE for all 1000 of the samples. In contrast to the results of Table 3b, the MCVME converged 1000 times out of 1000 for each set of starting values in Table 3a, while of course, in this situation the ML estimators  $\bar{X}$  and  $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  exist in closed form.

#### 4. Concluding Remarks

In this report we have briefly considered the use of the MHDE for estimating the parameters of the mixture of normals model. The MHDE was of interest originally due to its theoretical robustness and asymptotic full efficiency. Our empirical results indicate that these properties do hold in the mixture setting, at least to some degree. We have shown that the MHDE requires computation times which are similar to those for the other techniques although it is more difficult to calculate. Further research is in progress concerning the use of density estimators other than the Epanechnikov kernel density estimator. Preliminary results indicate that MHD estimates based upon the histogram density estimator (Tapia and Thompson(1978)) require substantially less computer time than those based on the Epanechnikov kernel, and they have only slightly higher MSEs.

The major problem concerning the use of the MHDE appears to be with its extreme sensitivity to starting values. It is our opinion that, although, these convergence problems could be somewhat alleviated with further "fine-tuning" of the iterative algorithm, the implementation of the MHDE into segment level proportion estimation procedures would be difficult.

# Appendix

Let  $f(x) = pf_1(x) + (1-p)f_2(x)$  where

$$f_i(x) = \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2}$$

and let  $\underline{\theta} = (\mu_1, \sigma_1, \mu_2, \sigma_2, p)'$ . Then

$$\frac{\partial f^{1/2}}{\partial \underline{\theta}} = \left( \frac{\partial f^{1/2}}{\partial \mu_1}, \frac{\partial f^{1/2}}{\partial \sigma_1}, \frac{\partial f^{1/2}}{\partial \mu_2}, \frac{\partial f^{1/2}}{\partial \sigma_2}, \frac{\partial f^{1/2}}{\partial p} \right),$$

where

$$\frac{\partial f^{1/2}}{\partial \mu_1} = \frac{p}{2} f^{-1/2} \frac{\partial f_1}{\partial \mu_1}$$

$$\frac{\partial f^{1/2}}{\partial \sigma_1} = \frac{p}{2} f^{-1/2} \frac{\partial f_1}{\partial \sigma_1}$$

$$\frac{\partial f^{1/2}}{\partial \mu_2} = \frac{(1-p)}{2} f^{-1/2} \frac{\partial f_2}{\partial \mu_2}$$

$$\frac{\partial f^{1/2}}{\partial \sigma_2} = \frac{(1-p)}{2} f^{-1/2} \frac{\partial f_2}{\partial \sigma_2}$$

$$\frac{\partial f^{1/2}}{\partial p} = \frac{1}{2} f^{-1/2} (f_1 - f_2)$$

and

$$\frac{\partial f_i}{\partial \mu_i} = \frac{1}{\sqrt{2\pi} \sigma_i^2} \left( \frac{x-\mu_i}{\sigma_i} \right) e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2}, \quad i = 1, 2$$

$$\frac{\partial f_i}{\partial \sigma_i} = \frac{1}{\sqrt{2\pi} \sigma_i^2} \left[ \left( \frac{x - \mu_i}{\sigma_i} \right)^2 - 1 \right] e^{-\frac{1}{2} \left( \frac{x - \mu_i}{\sigma_i} \right)^2}, \quad i = 1, 2.$$

Similarly,

$$\frac{\partial^2 f^{1/2}}{\partial \theta^2} = \begin{pmatrix} \frac{\partial^2 f^{1/2}}{\partial \mu_1^2} & \frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial \sigma_1} & \frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial \mu_2} & \frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial \sigma_2} & \frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial p} \\ \frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial \sigma_1} & \frac{\partial^2 f^{1/2}}{\partial \sigma_1^2} & \frac{\partial^2 f^{1/2}}{\partial \sigma_1 \partial \mu_2} & \frac{\partial^2 f^{1/2}}{\partial \sigma_1 \partial \sigma_2} & \frac{\partial^2 f^{1/2}}{\partial \sigma_1 \partial p} \\ \frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial \mu_2} & \frac{\partial^2 f^{1/2}}{\partial \sigma_1 \partial \mu_2} & \frac{\partial^2 f^{1/2}}{\partial \mu_2^2} & \frac{\partial^2 f^{1/2}}{\partial \mu_2 \partial \sigma_2} & \frac{\partial^2 f^{1/2}}{\partial \sigma_2 \partial p} \\ \frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial \sigma_2} & \frac{\partial^2 f^{1/2}}{\partial \sigma_1 \partial \sigma_2} & \frac{\partial^2 f^{1/2}}{\partial \mu_2 \partial \sigma_2} & \frac{\partial^2 f^{1/2}}{\partial \sigma_2^2} & \frac{\partial^2 f^{1/2}}{\partial \sigma_2 \partial p} \\ \frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial p} & \frac{\partial^2 f^{1/2}}{\partial \sigma_1 \partial p} & \frac{\partial^2 f^{1/2}}{\partial \mu_2 \partial p} & \frac{\partial^2 f^{1/2}}{\partial \sigma_2 \partial p} & \frac{\partial^2 f^{1/2}}{\partial p^2} \end{pmatrix}$$

where

$$\frac{\partial^2 f^{1/2}}{\partial \mu_1^2} = \frac{-p^2}{4} f^{-3/2} \left( \frac{\partial f_1}{\partial \mu_1} \right)^2 + \frac{p}{2} f^{-1/2} \left( \frac{\partial^2 f_1}{\partial \mu_1^2} \right)$$

$$\frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial \sigma_1} = \frac{-p^2}{4} f^{-3/2} \left( \frac{\partial f_1}{\partial \mu_1} \right) \left( \frac{\partial f_1}{\partial \sigma_1} \right) + \frac{p}{2} f^{-1/2} \left( \frac{\partial^2 f_1}{\partial \mu_1 \partial \sigma_1} \right)$$

$$\frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial \mu_2} = \frac{-p(1-p)}{4} f^{-3/2} \left( \frac{\partial f_1}{\partial \mu_1} \right) \left( \frac{\partial f_2}{\partial \mu_2} \right)$$



$$\frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial \sigma_2} = \frac{-p(1-p)}{4} f^{-3/2} \left( \frac{\partial f_1}{\partial \mu_1} \right) \left( \frac{\partial f_2}{\partial \sigma_2} \right)$$

$$\frac{\partial^2 f^{1/2}}{\partial \mu_1 \partial p} = \frac{1}{2} f^{-1/2} \left( \frac{\partial f_1}{\partial \mu_1} \right) - \frac{p}{4} f^{-3/2} \left( \frac{\partial f_1}{\partial \mu_1} \right) (f_1 - f_2)$$

$$\frac{\partial^2 f^{1/2}}{\partial \sigma_1^2} = \frac{-p^2}{4} f^{-3/2} \left( \frac{\partial f_1}{\partial \sigma_1} \right)^2 + \frac{p}{2} f^{-1/2} \left( \frac{\partial^2 f_1}{\partial \sigma_1^2} \right)$$

$$\frac{\partial^2 f^{1/2}}{\partial \mu_2 \partial \sigma_1} = \frac{-p(1-p)}{4} f^{-3/2} \left( \frac{\partial f_1}{\partial \sigma_1} \right) \left( \frac{\partial f_2}{\partial \mu_2} \right)$$

$$\frac{\partial^2 f^{1/2}}{\partial \sigma_1 \partial \sigma_2} = \frac{-p(1-p)}{4} f^{-3/2} \left( \frac{\partial f_1}{\partial \sigma_1} \right) \left( \frac{\partial f_2}{\partial \sigma_2} \right)$$

$$\frac{\partial^2 f^{1/2}}{\partial \sigma_1 \partial p} = \frac{1}{2} \left( \frac{\partial f_1}{\partial \sigma_1} \right) [f^{-1/2} - \frac{p}{2} f^{-3/2} (f_1 - f_2)]$$

$$\frac{\partial^2 f^{1/2}}{\partial \mu_2^2} = \frac{-(1-p)^2}{4} f^{-3/2} \left( \frac{\partial f_2}{\partial \mu_2} \right)^2 + \frac{(1-p)}{2} f^{-1/2} \left( \frac{\partial^2 f_2}{\partial \mu_2^2} \right)$$

$$\frac{\partial^2 f^{1/2}}{\partial \mu_2 \partial \sigma_2} = \frac{-(1-p)^2}{4} f^{-3/2} \left( \frac{\partial f_2}{\partial \mu_2} \right) \left( \frac{\partial f_2}{\partial \sigma_2} \right) + \frac{(1-p)}{2} f^{-1/2} \left( \frac{\partial^2 f_2}{\partial \mu_2 \partial \sigma_2} \right)$$

$$\frac{\partial^2 f^{1/2}}{\partial \mu_2 \partial p} = -\frac{1}{2} f^{-1/2} \left( \frac{\partial f_2}{\partial \mu_2} \right) - \frac{(1-p)}{4} f^{-3/2} \left( \frac{\partial f_2}{\partial \mu_2} \right) (f_1 - f_2)$$

$$\frac{\partial^2 f^{1/2}}{\partial \sigma_2^2} = \frac{-(1-p)^2}{4} f^{-3/2} \left( \frac{\partial f_2}{\partial \sigma_2} \right)^2 + \frac{(1-p)}{2} f^{-1/2} \left( \frac{\partial^2 f_2}{\partial \sigma_2^2} \right)$$

$$\frac{\partial^2 f^{1/2}}{\partial \sigma_2 \partial p} = \frac{1}{2} \left( \frac{\partial f_2}{\partial \sigma_2} \right) [-f^{-1/2} - \frac{(1-p)}{2} f^{-3/2} (f_1 - f_2)]$$

$$\frac{\partial^2 f^{1/2}}{\partial p^2} = -\frac{1}{4} f^{-3/2} (f_1 - f_2)^2$$

and

$$\frac{\partial^2 f_i}{\partial \mu_i^2} = \frac{1}{\sqrt{2\pi} \sigma_i^3} \left[ \left( \frac{x-\mu_i}{\sigma_i} \right)^2 - 1 \right] e^{-\frac{1}{2} \left( \frac{x-\mu_i}{\sigma_i} \right)^2}, \quad i = 1, 2$$

$$\frac{\partial^2 f_i}{\partial \sigma_i^2} = \frac{1}{\sqrt{2\pi} \sigma_i^3} \left[ 2 - 5 \left( \frac{x-\mu_i}{\sigma_i} \right)^2 + \left( \frac{x-\mu_i}{\sigma_i} \right)^4 \right] e^{-\frac{1}{2} \left( \frac{x-\mu_i}{\sigma_i} \right)^2}, \quad i = 1, 2$$

$$\frac{\partial^2 f_i}{\partial \mu_i \partial \sigma_i} = \frac{1}{\sqrt{2\pi} \sigma_i^3} \left[ \left( \frac{x-\mu_i}{\sigma_i} \right)^3 - 3 \left( \frac{x-\mu_i}{\sigma_i} \right) \right] e^{-\frac{1}{2} \left( \frac{x-\mu_i}{\sigma_i} \right)^2}, \quad i = 1, 2.$$

## REFERENCES

1. Beran, R. (1977). "Minimum Hellinger Distance Estimates for Parametric Models," Annals of Statistics 5, 445-463.
2. Bickel, P.J. (1978). "Some Recent Developments in Robust Statistics," Paper presented to the Fourth Australian Statistical Conference.
3. Stather, C. R. (1981). "Robust Statistical Inference Using Hellinger Distance Methods," Ph.D. dissertation, La Trobe University, Australia.
4. Tapia, R.A. and Thompson, J.R. (1978). Nonparametric Probability Density Estimation. Baltimore: Johns Hopkins Univ. Press.
5. Woodward, W.A., Parr, W.C., Schucany, W.R., and Gray, H.L. (1983). "Minimum Distance Estimation of Mixture Model Parameters - Asymptotic Results and Simulation Comparisons with Maximum Likelihood," NASA Technical Report SR-63-04427, June 1983.
6. Woodward, W.A., Schucany, W.R., Lindsey, H. and Gray, H.L. (1982). "A Comparison of Minimum Distance and Maximum likelihood Techniques for Proportion Estimation," NASA technical report SR-62-04376, November 1982.

**APPENDIX C**

**Proportion Estimation in Mixtures of  
Asymmetric Distributions**

**(SR-63-04409)**

# PROPORTION ESTIMATION IN MIXTURES OF ASYMMETRIC DISTRIBUTIONS

Wayne A. Woodward, Richard F. Gunst,  
Hildegard Lindsey, and H. L. Gray

Center for Applied Mathematical and Statistical Research  
Southern Methodist University

## 1. Introduction

A standard approach to the estimation of crop proportions in agricultural remote sensing has been to estimate the proportions  $p_1, p_2, \dots, p_m$  in the mixture density

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + \dots + p_m f_m(x) \quad (1.1)$$

where  $m$  is the number of components (crops) in the mixture and  $f_i(x)$  is the density associated with component  $i$ . The usual procedure for estimating the parameters in the mixture model of (1.1) has been to:

- (a) assume that the component distributions are normal
- (b) use maximum likelihood estimation.

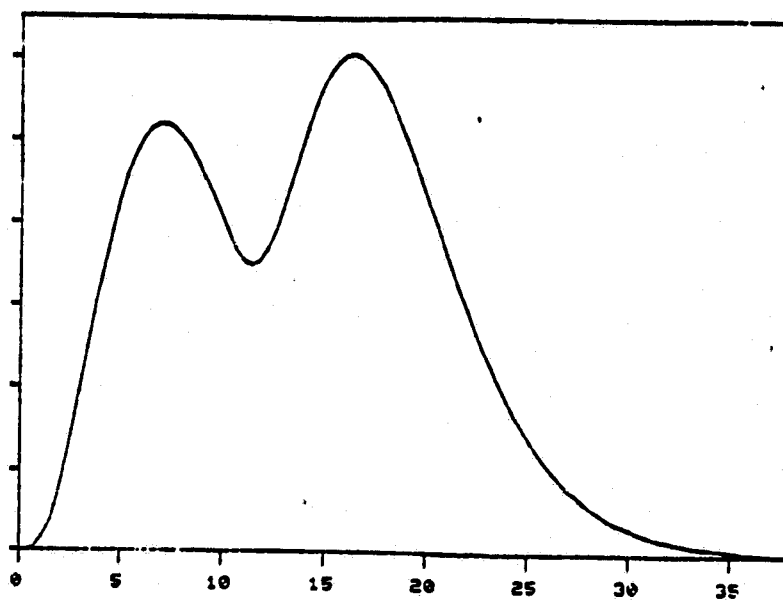
The variable  $X$  has usually been taken to be the reflected energy in the four LANDSAT bands or some linear combination of these such as greenness or brightness. Recent efforts have focused on the use of certain derived features from growth models such as  $g_{\max}$  and  $t_{\max}$  as variables in the mixture model. Studies have indicated that there is often a

substantial asymmetry in the distributions of these features for a given crop. Woodward et. al.(1982) have shown that asymmetry in the component distributions can cause a substantial bias in the proportion estimators when the mixture of normals model is assumed. As an example, in Figure 1 we display the mixture density associated with the mixture of two distributions. Examination of the figure reveals that if the component distributions are assumed to be symmetric, then we must conclude that  $p_1 < p_2$  and that the component to the right has larger variance. Actually, in this mixture  $p_1 = p_2$  and the distribution to the left is a  $\chi^2(9)$  while the component to the right is a "shifted"  $\chi^2(9)$ , i.e. its left truncation point is at  $x=10$  instead of  $x=0$ . It can be seen that a bias will be introduced in estimating mixing proportions in this mixture if the component distributions are assumed to be symmetric, which of course is the case when the components are assumed to be normal.

In this paper we will discuss techniques for estimating the crop proportions in the presence of asymmetric component distributions. In particular the estimation procedures we will propose assume that the underlying component distributions belong to some family of distributions whose members can be either symmetric or skewed depending on parameter configurations. At the present time, the Weibull distribution is being examined concerning its usefulness in this area. The effectiveness of this technique will be examined through simulations.

FIGURE 1

A Mixture Density



## 2. The Weibull Distribution

The Weibull distribution is named after the Swedish physicist Waloddi Weibull who used it to represent the distribution of the breaking strength of materials (Weibull(1939)). The distribution has been widely used in recent years in the fields of reliability and quality control. Its popularity is largely due to the flexibility which it introduces into the model due to the fact that it can be used to describe distributions which are symmetric or skewed in either direction. For these reasons we have chosen to investigate its applicability to estimation in mixtures of asymmetric components. The three-parameter Weibull density can be expressed as

$$f(x) = \frac{\gamma}{\beta} \left(\frac{x-\alpha}{\beta}\right)^{\gamma-1} e^{-\left(\frac{x-\alpha}{\beta}\right)^{\gamma}}, \quad x \geq \alpha \quad (2.1)$$

$\beta, \gamma > 0$

We will use the notation  $X \sim W(a,b,c)$  to indicate that the random variable  $X$  has a three-parameter Weibull distribution with parameters  $\alpha=a$ ,  $\beta=b$ , and  $\gamma=c$ . The parameter  $\alpha$  locates the left truncation point and  $\beta$  serves as a scale parameter while  $\gamma$  determines the shape of the distribution. In Figure 2 we show Weibull densities for a fixed  $\alpha$  and  $\beta$  and a range of values for  $\gamma$ . From the figure it is clear that the shape can vary dramatically as  $\gamma$  changes. In Figure 3 the



FIGURE 2

Weibull Densities with  $\alpha = 0$ ,  $\beta = 1$ ,  
and Various Values for  $\gamma$

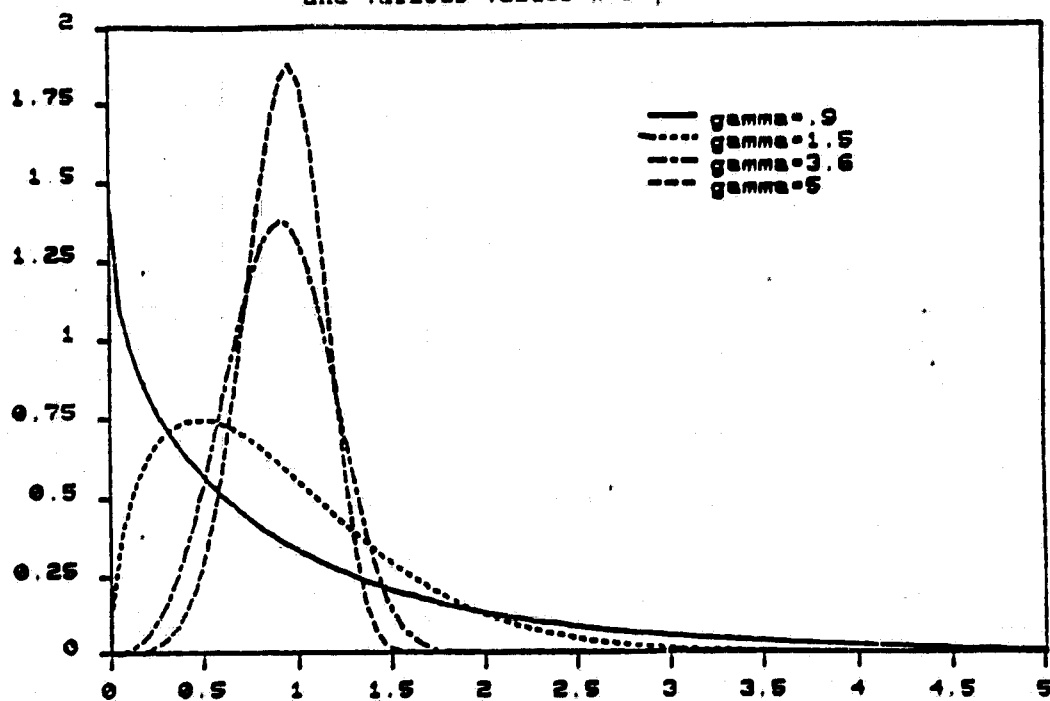
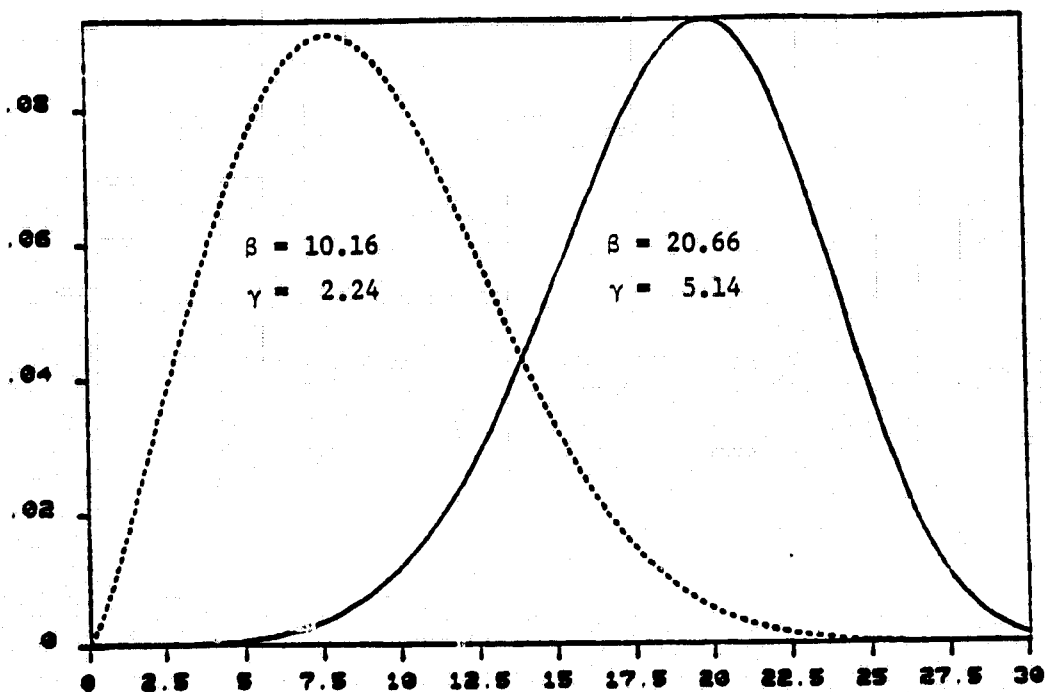


FIGURE 3

Weibull Densities with  $\alpha=0$



fact that the Weibull density can be skewed to the left as well as to the right is more clearly demonstrated. For  $\gamma=3.60232$  approximately, the standardized skewness parameter  $\beta_1 = \frac{\mu_3}{\mu_1^{3/2}}$ , where  $\mu_i$  is the  $i$ th central moment, is zero indicating symmetry. If  $\gamma < 3.60232$  then the Weibull is skewed to the right, while if  $\gamma > 3.60232$  it is skewed to the left. The Weibull distribution is unimodal, and if  $\gamma > 1$  the mode occurs at

$$x_m = \alpha + \beta \left( \frac{\gamma-1}{\gamma} \right)^{1/\gamma}.$$

Otherwise, when  $0 < \gamma < 1$ , the mode occurs at  $x_m = \alpha$ .

Dubey (1967) has studied the Weibull distribution when  $\gamma=3.60232$  and has concluded that it is very similar to the normal. In particular, Dubey has shown that

$$\sup_{-3 \leq v \leq 3} |F_Z(v) - F_Y(v)| \quad (2.3)$$

where  $F_Z$  denotes the cumulative distribution function of the random variable  $Z \sim N(0,1)$  and  $Y$  is the standardized variate  $Y = (X - \mu)/\sigma$  where  $\mu$  and  $\sigma^2$  are the mean and variance of the Weibull variate  $X$ .

It should be noted that the Weibull distribution is often given in the literature in two parameter form in which  $\alpha$  is assumed to be known (and usually 0). However, unless otherwise specified, reference to the Weibull distribution in this report, we will be to the three-parameter form specified by (2.1).

The cumulative distribution function corresponding to

the three-parameter Weibull is given by the closed form expression

$$F_X(x) = 1 - e^{-\left(\frac{x-\alpha}{\beta}\right)^\gamma} \quad (2.4)$$

while the noncentral moments are given by

$$u_r' = \sum_{k=0}^r \binom{r}{k} \alpha^{r-k} \beta^k \Gamma\left(\frac{k}{\gamma} + 1\right) \quad (2.5)$$

From (2.5) it can be seen that

$$\begin{aligned} \mu &= \alpha + \beta \Gamma\left(\frac{1}{\gamma} + 1\right) \\ \sigma^2 &= \beta^2 \left\{ \Gamma\left(\frac{2}{\gamma} + 1\right) - \Gamma^2\left(\frac{1}{\gamma} + 1\right) \right\}. \end{aligned} \quad (2.6)$$

The first three moments of the Weibull distribution determine the values of  $\alpha$ ,  $\beta$ , and  $\gamma$ . The method of moment estimators can be obtained using these relationships, but unfortunately the estimators do not exist in a closed form. The log-likelihood function for a random sample of  $n$  observations from the Weibull distribution is

$$\ln(L) = n \ln \gamma - n \gamma \ln \beta + (\gamma - 1) \sum_{i=1}^n \ln(x_i - \alpha) - \frac{1}{\beta^\gamma} \sum_{i=1}^n (x_i - \alpha)^\gamma \quad (2.7)$$

Differentiating  $\ln(L)$  yields the following likelihood equations

$$-(\gamma - 1) \sum_{i=1}^n (x_i - \alpha)^{-1} + \frac{\gamma}{\beta^\gamma} \sum_{i=1}^n (x_i - \alpha)^{\gamma-1} = 0 \quad (2.8)$$

$$\beta = \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \alpha)^\gamma \right]^{1/\gamma} \quad (2.9)$$

$$\gamma = \left\{ \sum_{i=1}^n \left[ \ln\left(\frac{x_i - \alpha}{\beta}\right) \right] \left[ \left(\frac{x_i - \alpha}{\beta}\right)^\gamma - 1 \right] \right\}^{-1} \quad (2.10)$$

Let  $\hat{\alpha}$ ,  $\hat{\beta}$ , and  $\hat{\gamma}$  denote the estimators obtained from the simultaneous solution of equations (2.8) to (2.10). If  $0 < \hat{\alpha} \leq Y_1$ , where  $Y_i$  denotes the  $i$ th order statistic, these estimators are the maximum likelihood (ML) estimators for the three Weibull parameters. However, due to the restriction  $x \geq \alpha$  in (2.1), if  $\hat{\alpha} > Y_1$ , then the MLE of  $\alpha$  is taken to be  $Y_1$  and  $\beta$  and  $\alpha$  are estimated from (2.9) and (2.10). As in the case of method of moment estimators, the ML estimators do not have a closed form expression. For a general review of the literature on Weibull parameter estimation see Johnson and Kotz (1970).

### 3. Mixtures of Weibull Distributions

In order to examine the feasibility of using the Weibull as a model for the component distributions in the mixture model of (1.1), we will investigate the estimation of the parameters in the mixture of two Weibull distributions. This mixture density is given in (3.1)

$$f(x) = p \frac{\gamma_1}{\beta_1} \left(\frac{x-\alpha_1}{\beta_1}\right)^{\gamma_1-1} e^{-\left(\frac{x-\alpha_1}{\beta_1}\right)^{\gamma_1}} + (1-p) \frac{\gamma_2}{\beta_2} \left(\frac{x-\alpha_2}{\beta_2}\right)^{\gamma_2-1} e^{-\left(\frac{x-\alpha_2}{\beta_2}\right)^{\gamma_2}} \quad (3.1)$$

where the 7 parameters  $p$ ,  $\alpha_1$ ,  $\beta_1$ ,  $\gamma_1$ ,  $\alpha_2$ ,  $\beta_2$ , and  $\gamma_2$  are assumed to be unknown.

Previous research in this area includes that of Kao (1959), who proposed a graphical procedure for estimating the parameters in (3.1) when one of the location parameters is assumed to be known and equal to zero. The estimation of

the 6 remaining parameters is accomplished using a graphical procedure whose applicability to our problem seems to be limited although some of his estimation rules could be automated. Rider(1961) and Falls(1970) propose estimating the parameters of a mixture of two-parameter Weibulls using the method of moments. Falls' procedure involves estimating the mixing proportion  $p$  using a graphical procedure similar to that of Kao.

Maximum likelihood estimation of the parameters of (3.1) has been discussed by Looney and Bargmann(1982). The likelihood equations obtained by differentiating the log-likelihood function  $\ln(L)$

$$\ln(L) = \sum_{i=1}^n \{ \ln[p f_1(x_i) + (1-p) f_2(x_i)] \}$$

with respect to each of the 7 parameters yields the likelihood equations

$$(\gamma_j - 1) \sum_{i=1}^n f(j|x_i) (x_i - \alpha_j)^{-1} - \frac{\gamma_j}{\beta_j^{\gamma_j}} \sum_{i=1}^n f(j|x_i) (x_i - \alpha_j)^{\gamma_j - 1} = 0, j=1, 2 \quad (3.2)$$

$$\beta_j = \left\{ \left[ \sum_{i=1}^n (x_i - \alpha_j)^{\gamma_j} f(j|x_i) \right] / \sum_{i=1}^n f(j|x_i) \right\}^{1/\gamma_j}, j=1, 2 \quad (3.3)$$

$$\gamma_j = \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i - \alpha_j}{\beta_j} \right)^{\gamma_j - 1} \ln \left( \frac{x_i - \alpha_j}{\beta_j} \right) \right] / \sum_{i=1}^n f(j|x_i) \right\}^{-1} j=1, 2 \quad (3.4)$$

$$p = \frac{1}{n} \sum_{i=1}^n f(1|x_i) \quad (3.5)$$

where  $f(i|x) = p_i f_i(x)/f(x)$  with  $f_i(x)$  denoting the  $i$ th component density and  $f(x)$  the mixture density. Solving this set of equations for the maximum likelihood estimators is difficult due largely to equations (3.2) which are not in fixed point form. Looney and Bargmann(1982) suggested a

procedure in which the shape parameters  $\gamma_1$  and  $\gamma_2$  are fixed independently at each of the values

$$\{\frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1, \frac{3}{2}, 2, 3, 4, 5\}$$

and, for each of the  $(\hat{\gamma}_1, \hat{\gamma}_2)$  pairs, "preliminary" maximum likelihood estimates of the remaining 5 parameters are found. A search procedure results in selecting the  $(\hat{\gamma}_1, \hat{\gamma}_2)$  pair for which  $\ln(L)$  is maximized. With  $\hat{\gamma}_1$  and  $\hat{\gamma}_2$  fixed at these values, maximum likelihood estimation for the remaining 5 parameters is then carried through to convergence. The Looney and Bargmann procedure for solving the system of equations (3.2) - (3.5) seems overly restrictive with respect to the selection of possible values of the shape parameter, while expansion of the search procedure to allow for more shape parameter values would probably be prohibitive because of time constraints. However, solution of these likelihood equations directly appears to us to be quite intractable. For these reasons, we have investigated the use of minimum distance(MD) estimation, first introduced by Wolfowitz(1957), for estimating the 7 parameters in the mixture of Weibulls model given in (3.1). Woodward et. al.(1982) have recently studied the use of MD estimation in the mixture of normals model. These authors showed that MD estimation was easy to implement in that setting, and that MD estimators showed to be superior to ML estimators under departures from component normality. Since our use of Weibull components is due to the

flexibility which it introduces into the model rather than underlying theoretical justifications, we definitely need an estimation procedure which is robust to departures from assumptions.

The minimum distance estimator of the parameter  $\theta$  (possibly vector valued) is defined to be that value of  $\theta$  which minimizes the distance between  $H_\theta$  and  $F_n$  where  $H = \{H_\theta : \theta \in \Omega\}$  denotes a family of distributions depending on  $\theta$  and  $F_n$  denotes the empirical distribution function, i.e.  $F_n(x) = k/n$  where  $k$  is the number of observations less than or equal to  $x$ . The family of distributions  $H$  is referred to as the projection model, where in this case  $\theta = (p, \alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2)$ , and  $H_\theta(x)$  is the distribution function associated with a mixture of two Weibull components given by

$$H_\theta(x) = p \left[ 1 - e^{-\left(\frac{x-\alpha_1}{\beta_1}\right)^{\gamma_1}} \right] + (1-p) \left[ 1 - e^{-\left(\frac{x-\alpha_2}{\beta_2}\right)^{\gamma_2}} \right]. \quad (3.6)$$

Note that in contrast to the situation in which the projection model is taken to be the mixture of two normals,  $H_\theta(x)$  in (3.6) has a closed form expression. The choice of distance function to be used to measure the distance between two distributions is a topic of current interest in the field of MD estimation. Woodward et. al. (1982) used the Cramér-von Mises distance,  $W^2$ , given by

$$W^2 = \int_{-\infty}^{\infty} [G_1(x) - G_2(x)]^2 dG_2(x) \quad (3.7)$$

where  $G_1$  and  $G_2$  are two distribution functions, and we have

chosen to use this distance measure in the current study. The distance between a distribution function  $H_0$  and the empirical distribution function  $F_n$ , which is needed for calculation of the MD estimator, is given by the simplified expression

$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^n [H_0(Y_i) - \frac{i-.5}{n}]^2, \quad (3.8)$$

where  $Y_i$  denotes the  $i$ th order statistic. Since  $H_0(X)$  exists in closed form, the MDE in this case is easily obtained by using nonlinear least squares techniques to minimize (3.8). We have performed this minimization with IMSL subroutine ZXSSQ which uses Marquardt's(1963) procedure.

#### 4. Simulation Results

In Section 3 we discussed the problem of estimation in the mixture of Weibulls model. From that discussion it appears that the minimum distance techniques are preferable for estimating the parameters in a mixture of three parameter Weibulls, especially in terms of computational convenience. In this section we will discuss the results of an initial computer simulation which was designed for use in evaluating the numerical capabilities of this method. All computations were performed on the CDC 6600 at Southern Methodist University. In this section we will evaluate the performance of the MD estimation procedures discussed. Since the usual procedure is to assume that the components are normal, we will compare the Weibull based MDEs with the



normal based procedures. We have generated samples from mixtures of normal components and mixtures of  $\chi^2(9)$  components. Obviously, we would expect the normal based procedures to perform better than Weibull based procedures when the mixture really is a mixture of normal components. However, if the Weibull techniques are to be useful, then they must give reasonable results in this situation since the normal assumption does appear to be a reasonable assumption in some cases. Since the Weibull with  $\gamma=3.6$  is very nearly normal, there is reason to believe that Weibull procedures will perform well in this situation. We have not simulated samples from mixtures of Weibull distributions, but we plan to consider this in the future. Of course, as mentioned in the previous section, we are most interested in the performance of the Weibull based procedures when the underlying components from which we sample are not necessarily Weibulls, but are realistic representatives of the types of component distributions we see in practice.

Our simulation results are based on 200 samples of size  $n=200$  from mixtures of normal and of  $\chi^2(9)$  components. In each mixture, the variance associated with the two components are equal. In fact, the two component distributions differ from each other only by a location shift. We have simulated from mixtures having mixing proportions of .25, .50, and .75, and with varying degrees of separation between the two component distributions. Overlap as defined by Woodward et.al.(1982) is a

quantification of this separation. It is defined as the the probability of misclassification using the rule:

Classify an observation  $x$  as:

population 1 if  $x < x_c$

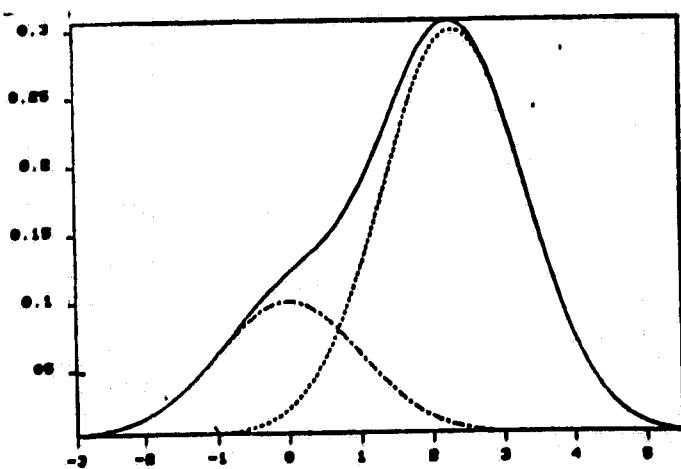
population 2 if  $x \geq x_c$

where without loss of generality, population 1 is assumed to be centered to the left of population, and where  $x_c$  is the unique point between  $\mu_1$  and  $\mu_2$  such that

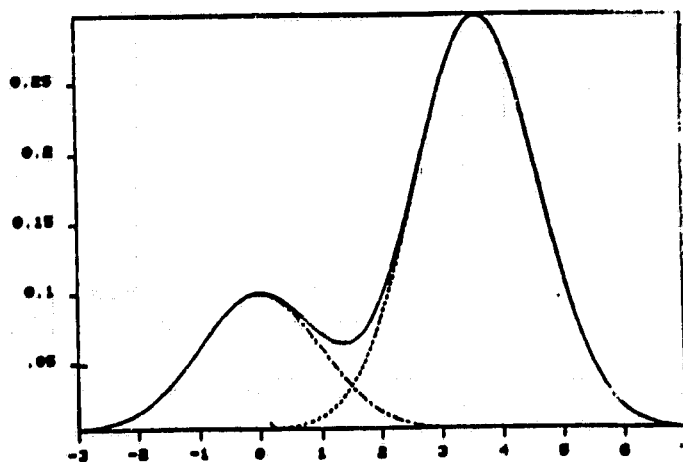
$$pf_1(x_c) = (1-p)f_2(x_c).$$

We have based our current study on "overlaps" of .03 and .10. In Figure 4 we display the mixture densities associated with normal components. For each mixture, the scaled components  $pf_1(x)$  and  $(1-p)f_2(x)$  are also shown. Note that the densities for  $p=.75$  are not displayed here. Since  $\sigma_1=\sigma_2$ , it follows that  $f^p(x)=f^{1-p}(\mu_1+\mu_2-x)$  where  $f^p(x)$  denotes the mixture density associated with a mixing proportion of  $h$ . Thus the shapes of the densities at  $p=.75$  can be inferred from those at  $p=.25$ . Likewise, parameter estimation for  $p=.75$  is not included in the results of the simulations for the mixtures of normals. In Figure 5 we display the mixture densities associated with the mixtures of  $\chi^2(9)$  components. Note that although we refer to a mixture of  $\chi^2(9)$

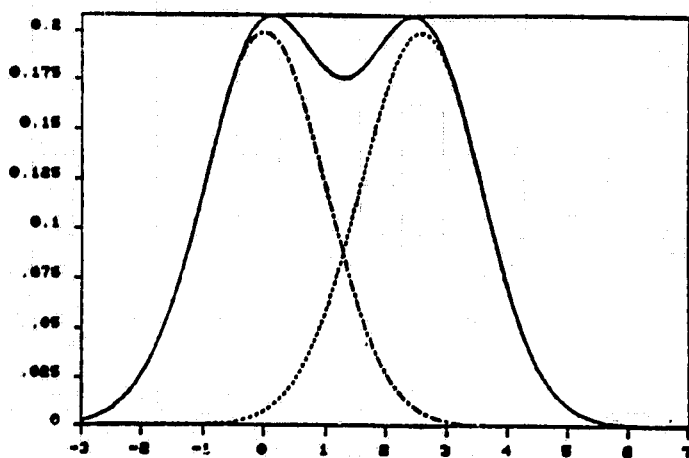
FIGURE 4  
Mixture Densities with Normal Components



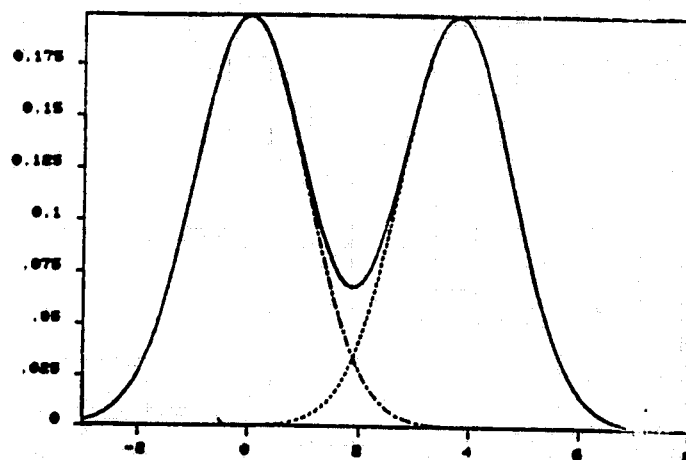
(a)  $p = .25$ , Overlap = .10



(b)  $p = .25$ , Overlap = .03

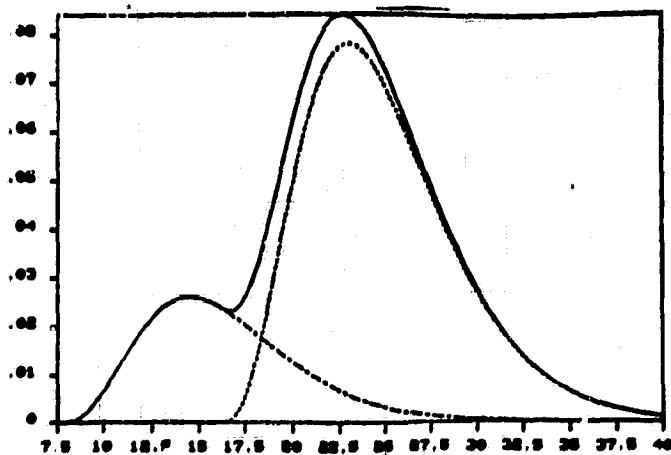


(c)  $p = .50$ , Overlap = .10

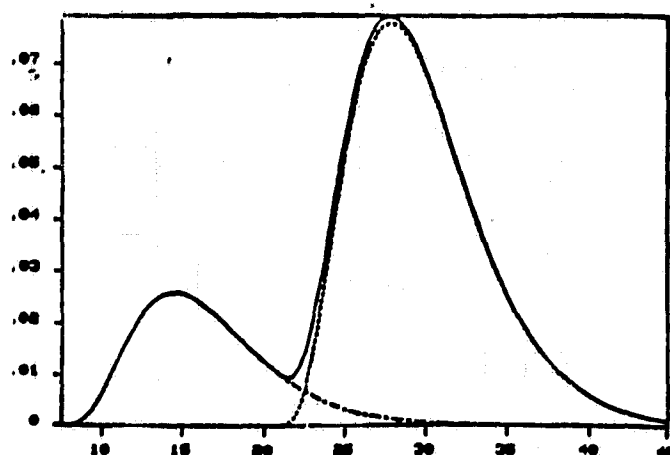


(d)  $p = .50$ , Overlap = .03

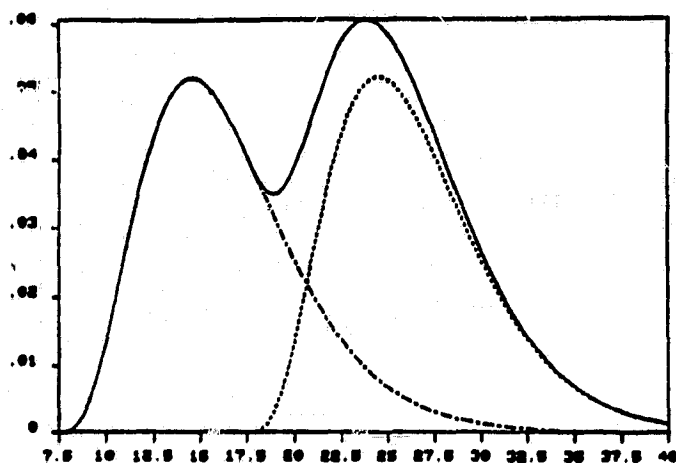
FIGURE 5  
Mixtures Densities with  $\chi^2(9)$  Components



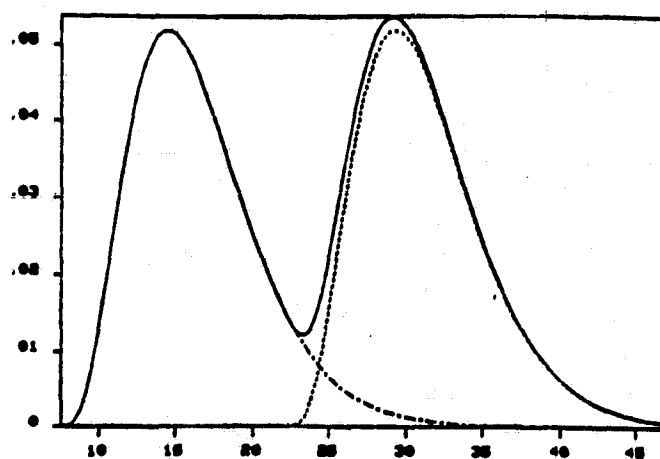
(a)  $p = .25$ , Overlap = .10



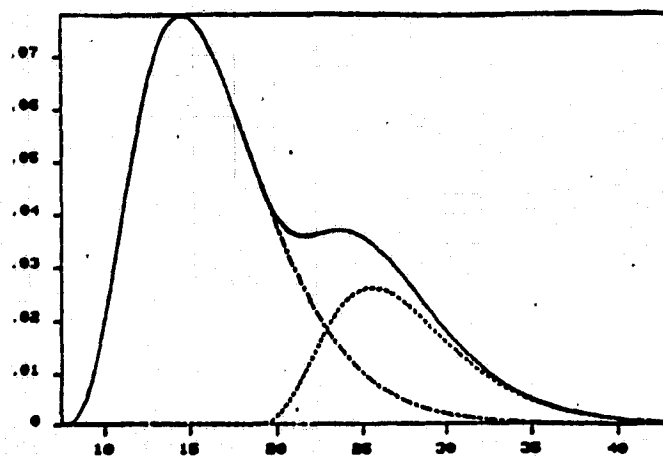
(b)  $p = .25$ , Overlap = .03



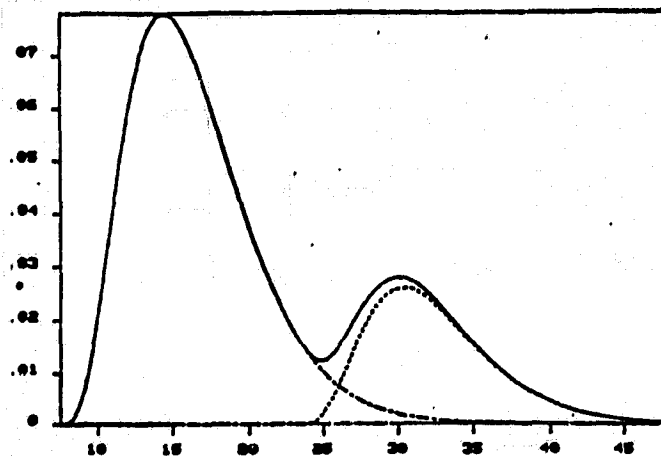
(c)  $p = .50$ , Overlap = .10



(d)  $p = .50$ , Overlap = .03



(e)  $p = .75$ , Overlap = .10



(f)  $p = .75$ , Overlap = .03

distributions here, they are actually "shifted" chi-squares, i.e. the left truncation points are different from zero.

For each of the simulated samples, three sets of parameter estimates were obtained:

- (1) ML estimates based on mixture of normals model (MLEN)
- (2) MD estimates based on mixture of normals model (MDEN)
- (3) MD estimates based on mixture of Weibulls model (MDEW)

Although the MLEN and MDEN provide estimates of all 5 of the parameters of the mixture of normals model, and the MDEW produces estimates for all 7 parameters in the mixture of Weibulls model, only the results for the estimation of  $p$  will be shown. The mixing proportion is the parameter of primary interest, and when dealing with the "wrong-model" situations, the remaining parameter estimates often do not have a meaningful interpretation. For purposes of aiding in the discussions which follow, we will call a component model from which we actually simulated, a "simulation component model", while a component model which is assumed under a particular estimation procedure will be called an "estimation component model". Thus, a "wrong-model" situation is one in which the simulation component models are not the same as the estimation component models.

In the "correct-model" situations, i.e. using the MLEN or MDEN to estimate the parameters of a simulated mixture of normal components, the true parameter values are used as starting values for the iterative estimation procedures. In all of the other cases, there is not a "true" set of

parameters. For starting values, we have used the "true" mixing proportion, and then estimated the parameters of each component separately using a method of moments procedure. Consider a situation in which the estimation components are normal. We obtain starting values for each component by equating the first and second moments of the corresponding simulation and estimation components and using these to obtain  $\mu_1$  and  $\sigma_1^2$  for the normal estimation component. When the estimation components are Weibull, we have taken the approach of setting the starting value for  $\gamma$  at  $\gamma = 3.6$  for each component. Then the first two moments of the corresponding simulation and estimation components are equated to yield starting value estimates for the other two parameters. We believe that this provides a "neutral" start. If the final estimates reflect the finding of substantial skewness for one or both of the component Weibulls, this will be because of the data and not because of "skewed" starting values.

The normal component models were generated with  $\mu_1 = 7.5$ ,  $\sigma_1^2 = \sigma_2^2 = 1$ , and  $\mu_2$  positioned so that the desired overlap is obtained. As mentioned previously, both components in the chi-square mixtures were "shifted" chi-squares. In our simulations, the left truncation point for population 1 was always taken to be 7.5, and for population 2 it was located so that the desired overlap was obtained. In the MLEN and MDEN procedures, the natural constraints  $\sigma_1^2 > 0$ ,  $\sigma_2^2 > 0$ , and  $0 < p < 1$  were imposed. Similarly, for the MDEW, the natural

constraints  $\beta_1 > 0$ ,  $\gamma_1 > 0$ ,  $\beta_2 > 0$ ,  $\gamma_2 > 0$ , and  $0 \leq p \leq 1$  were imposed along with the constraints  $\alpha_1 \geq 0$  and  $\alpha_2 \geq 0$  which are reasonable constraints on the left-truncation point which would be imposed due to physical considerations, etc.

In Table 1 we display the results of the simulations. For a given simulation model and estimation procedure, we will obtain an estimate  $\hat{p}$  of  $p$ , defined by

$$\hat{p} = \frac{1}{n_s} \sum_{i=1}^{n_s} \hat{p}_i$$

where  $\hat{p}_i$  is the estimate of  $p$  for the  $i$ th sample, and  $n_s$  is the number of samples. Then based upon the simulations, estimates of the bias and MSE are given by:

$$\text{bias} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{p}_i - p) = \hat{p} - p$$

$$\text{MSE} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{p}_i - p)^2.$$

Upon viewing the results, it can be seen that the MDEW was competitive when the component models were actually normally distributed, and it produced the best overall results for the chi-square mixtures. Of particular interest is the chi-square mixture where  $p=.5$  and  $\text{overlap}=.10$ . This is the mixture displayed in Figure 5c and also in Figure 1 (except for location shift). When symmetric components are assumed (as with the MLEN and MDEN), a bias does occur in the estimation of  $p$  as discussed in Section 1. This behavior has been noted previously by Woodward, et.al.(1982). We see from the table that the MDEW performs substantially better

Table 1 - Simulation Results  
 Comparing Normal Based with  
 Weibull Based Estimation Procedures

Sample size = 200  
 Number of repetitions = 200

Mixture of Normals

		Overlap = .10			Overlap = .03		
		$\hat{\rho}$	Bias	MSE	$\hat{\rho}$	Bias	MSE
p = .25	MLEN	.27	.02	.022	.25	.00	.022
	MDEN	.37	.12	.074	.26	.01	.004
	MDEW	.34	.09	.044	.30	.05	.011
p = .5	MLEN	.50	.00	.014	.50	.00	.002
	MDEN	.49	-.01	.023	.47	-.03	.002
	MDEW	.48	-.02	.019	.51	.01	.004

Mixture of  $\chi^2(9)$

		Overlap = .10			Overlap = .03		
		$\hat{\rho}$	Bias	MSE	$\hat{\rho}$	Bias	MSE
p = .25	MLEN	.24	-.01	.061	.18	-.07	.006
	MDEN	.41	.16	.098	.17	-.08	.008
	MDEW	.50	.25	.122	.29	-.04	.007
p = .50	MLEN	.27	-.23	.064	.45	-.05	.011
	MDEN	.26	-.24	.061	.41	-.09	.010
	MDEW	.42	-.08	.024	.50	.00	.004
p = .75	MLEN	.50	-.25	.070	.65	-.10	.013
	MDEN	.48	-.27	.085	.64	-.11	.016
	MDEW	.62	-.13	.032	.71	.04	.005



than either of these normal based procedures on the basis of both bias and MSE. In Figure 6 we display histograms of the 200 estimates of  $p$  obtained from the three estimation procedures for the chi-square mixture shown in Figure 5c. It can be seen there, that the normal based procedures consistently estimated  $p$  to be substantially less than .5 while the estimates based on Weibull components were in general closer to the true value  $p=.5$ .

The one case in which the Weibull based estimates were not best, was when  $p=.25$  with  $\text{overlap}=.10$ . This mixture is displayed in Figure 5a where it is obvious that estimation should be difficult since there is no distinct contribution due to component 1 in the mixture. Indeed, all procedures yield poor estimates as measured by the high MSEs. In Figure 7, we display histograms of the  $p$  values obtained from the three estimation procedures for this set of parameter configurations. There it can be seen that the Weibull procedure certainly gave the poorest results, with estimates being spread nearly uniformly between 0 and 1. However, the normal based procedures also had difficulty as is reflected in the histograms. In fact, there appears to be a tendency for the  $\hat{p}_i$  values to be very low (approximately .10). However,  $\hat{p}$  is very close to .25 for the MLEN since several of the  $\hat{p}_i$  values were spread out uniformly between 0 and 1, which increased the estimate of  $p$  to near .25. However, the large MSE shown in the table for this case reflects this lack of accuracy.

FIGURE 6  
Histograms of Estimates of  $p$   
Based on 200 samples of size 200 from mixture shown in Figure 5(c)  
(True  $p = .5$ )

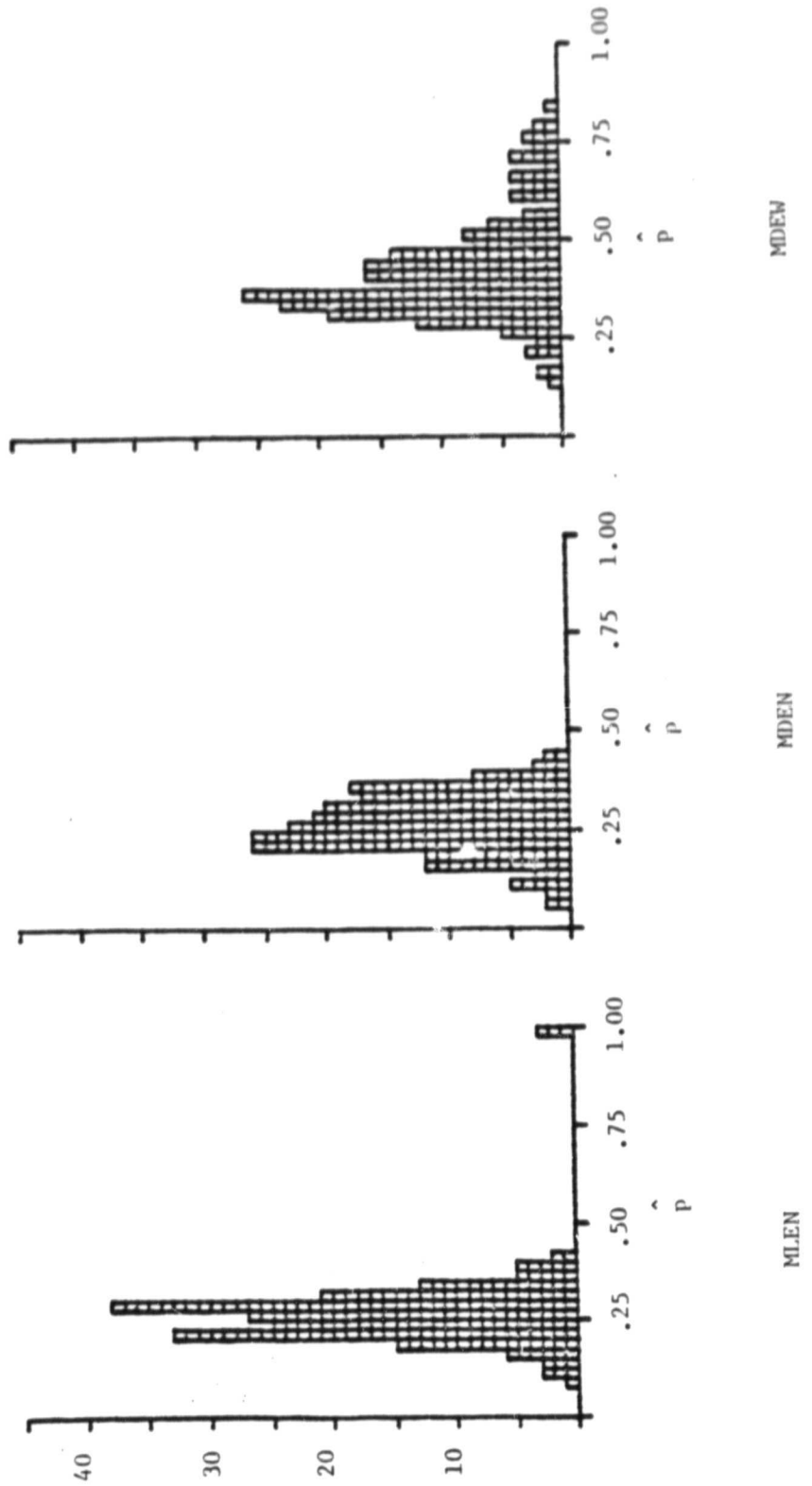
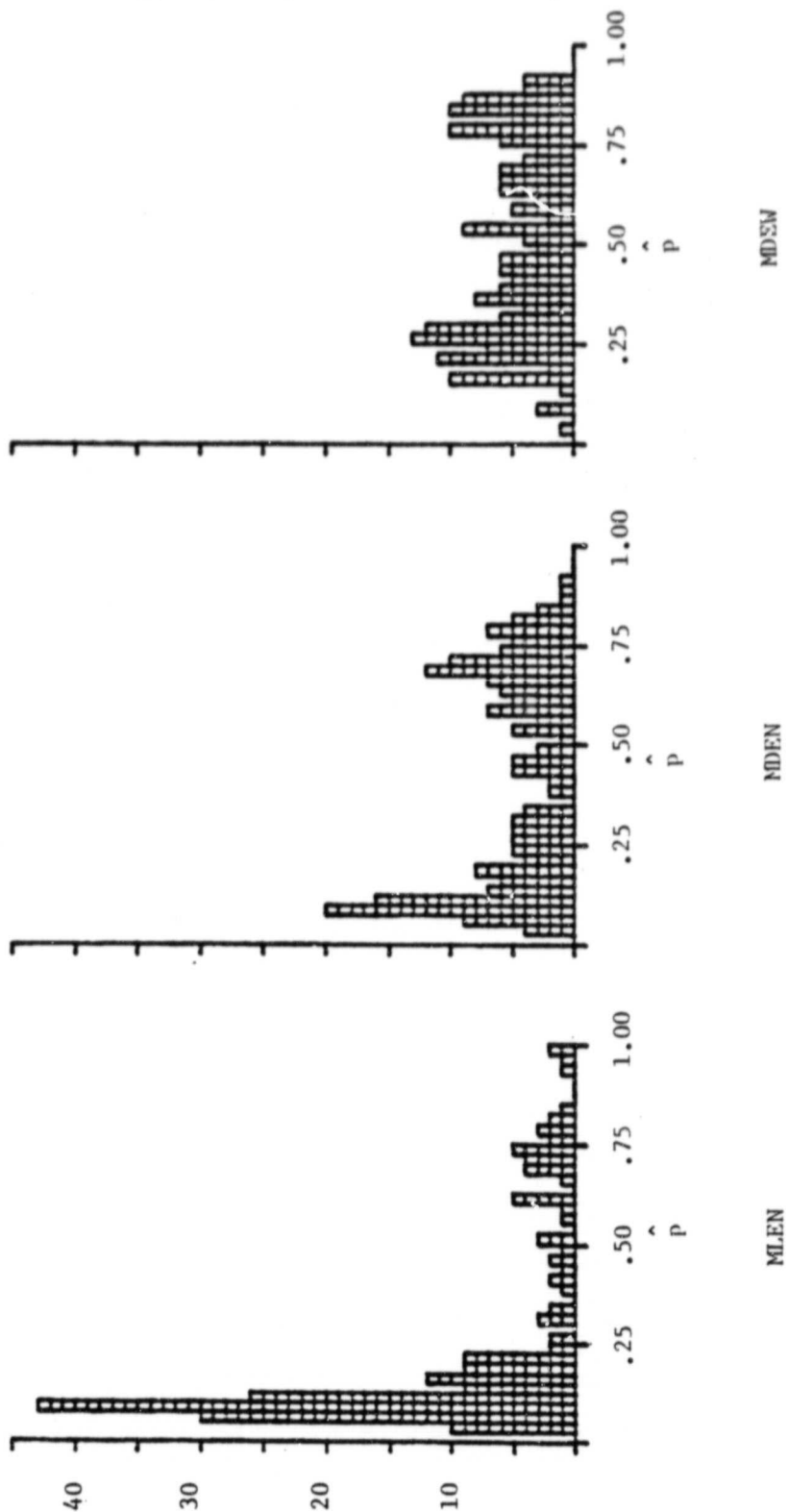


FIGURE 7

Histograms of Estimates of  $p$

Based on 200 samples of Size 200 from mixture shown in Figure 5(a)

(True  $p = .25$ )



ORIGINAL PAGE 19  
OF POOR QUALITY

## 5. Concluding Remarks

Results in this report and in the report by Woodward, et.al.(1982) indicate that the normal based procedures perform poorly in the presence of a mixture of asymmetric distributions. In this paper we have suggested the mixture of Weibulls model as an alternative to the mixture of normals model in this situation. Results indicate that minimum distance estimation of the parameters of a mixture of Weibulls is a viable alternative to the normal-based techniques currently in use.

Before this procedure could be recommended and implemented, further research is needed. For example, the problem of how to obtain starting values for the parameters of mixtures of possibly asymmetric components has not been resolved. Also, the Weibull based procedures should be applied to LANDSAT data in order to examine their performance on the types of asymmetry which will be encountered in practice. The fact that an additional parameter has been introduced into the model for each component has caused the estimation procedures to be slower than for the normal based procedures. Further investigation concerning the practical aspects of actually implementing the procedures is needed.

## References

1. Dubey, S. D. (1967). "Normal and Weibull Distributions," Naval Research Logistics Quarterly 14, 69-79.
2. Falls, L. W. (1970). "Estimation of Parameters in Compound Weibull Distributions," Technometrics 12, 399-407.
3. Kao, J. H. K. (1959). "A Graphical Estimation of Weibull Parameters in Life - testing of Electron Tubes," Technometrics 1, 389-407.
4. Johnson, N. L. and Kotz, S. (1970). Distributions in Statistics - Continuous Univariate Statistics - 2, Houghton Mifflin Company, Boston.
5. Looney, S. W. and Bargmann, R. E. (1982). "Procedures for Fitting a Mixture of Weibull Distributions to an Empirical Growth Curve," presented at National Meetings of the American Statistical Association in Cincinnati, August 1982.
6. Rider, P. R. (1961). "Estimating the Parameters of Mixed Poisson, Binomial, and Weibull Distributions by the Method of Moments," Bulletin de L'institute International de Statistique 39, 225-232.
7. Weibull, W. (1939). "A Statistical Theory of the Strength of Material," Ingeniors Vetenskaps Akademiens Handligar, Stockholm No. 151.
8. Wolfowitz, J. (1957). "The Minimum Distance Method," Annals of Mathematical Statistics 28, 75-88.

9. Woodward, W. A. , Schucany, W. R., Lindsey, H. and Gray, H. L. (1982). "A Comparison of Minimum Distance and Maximum Likelihood Techniques for Proportion Estimation," NASA Technical Report SR-62-04376, November 1982.